SOCIAL SCIENCES
PUBLIC LIBRARY
DETROIT

The JOURNAL of SOCIAL ISSUES

Vol. X No. 4

1954

Cross-National Research: A Case Study

Issue Editors: Eugene Jacobson and Stanley Schachter

Contents of Issue

	Page
Foreword ..... *Issue Editors*	2
Introduction ..... *Daniel Katz and Herbert Hyman*	5
Organizational Aspects of Cross-National Social Research *H. C. J. Duijker and Stein Rokkan*	8
A Comparative Study of Teachers' Attitudes to International Problems and Policies ..... *Vilhelm Aubert, Burton R. Fisher, and Stein Rokkan*	25
Methods Used for Producing Comparable Data in the OCSR Seven-Nation Attitude Study ..... *Eugene Jacobson*	40
Interpretative and Methodological Problems of Replicated Research *Stanley Schachter*	52
Notes on the Standardization of Experimental Manipulations and Measurements in Cross-National Research *Ragnar Rommetveit and Joachim Israel*	61
* * * * *	
Correspondence	69

Vol. X 1954 No. 4

THE JOURNAL OF SOCIAL ISSUES

Published Quarterly by the Association Press for

The SOCIETY FOR THE PSYCHOLOGICAL STUDY OF SOCIAL ISSUES

A Division of

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

Annual Subscriptions: $4.00 Single Copy $1.25

(For group rates and back numbers see last page)

STATEMENT OF PURPOSE

The Society for the Psychological Study of Social Issues is a group of nearly a thousand social psychologists and allied social scientists who share a concern with research on the psychlogical aspects of important social issues. SPSSI is governed by Kurt Lewin's dictum that "there is nothing so practical as a good theory." In various ways, the Society seeks to bring theory and practice into focus on human problems of the group, the community, and the nation as well as the increasingly important ones that have no national boundaries. This Journal has as its goal the communication of scientific findings and interpretations in a non-technical manner but without the sacrifice of professional standards.

Business address of publisher: Subscriptions, orders and business communications should be sent to ASSOCIATION PRESS, 291 Broadway, New York 7, N. Y. Subscribers (non-SPSSI members) should send changes in mailing address to Association Press.

Editorial address: Suggestions and inquiries concerning editorial content should be sent to: M. Brewster Smith, Social Science Research Council, 230 Park Avenue, New York 17, N. Y.

Society address: Dr. Alvin F. Zander, Secretary-Treasurer, Research Center for Group Dynamics, University of Michigan, Ann Arbor, Michigan.

Administrative Secretary: Requests for information about the Society and applications for membership should be sent to Helen S. Service, Institute for Social Research, University of Michigan, Ann Arbor, Michigan. SPSSI members should notify Mrs. Service about changes in mailing address.

Executive Editorial Committee: Richard Christie, Robert Harris, Eugene Hartley, Kenneth Herrold, Herbert Hyman, Brewster Smith, Leo Srole, David Truman.

Advisory Editorial Board: Floyd H. Allport, Gordon Allport, Hadley Cantril, Dorwin Cartwright, Stuart W. Cook, George W. Hartmann, Ernest R. Hilgard, Daniel Katz, Otto Klinberg, David Krech, Rensis Likert, Ronald Lippitt, Gardner Murphy, S. Stansfeld Sargent, Goodwin Watson.

Preface

International cooperation in the social sciences has been one of SPSSI's persisting interests. Early in the life of this *Journal,* an issue was devoted to progress and possibilities in international research ("Toward A Common Ground—International Social Science," J. S. Bruner, editor, Vol. 3, No. 1, 1947.) SPSSI has until the present been represented on the United States National Commission for UNESCO, and Otto Klineberg, former president and executive secretary, currently directs UNESCO's Department of Applied Social Science. Participating in the birth and early gropings of international social science organization, SPSSI has shared its ambitious goals and been tolerant of its sometimes faltering accomplishments.

Besides formal organization on the grand scale, there is another approach to international scientific cooperation: the collaboration of like-minded social scientists on particular concrete research undertakings. The Organization for Comparative Social Research, the work of which is discussed in the present issue, is one of the most notable examples of this promising new development. The Editor therefore takes much satisfaction in turning this issue over to an international team of collaborators on a project that was international through-and-through—in conception, planning, subject matter, methodology, and authorship.

The substantive results of the OCSR project, while they are touched on incidentally in this issue, are being reported elsewhere in detail. The present discussion focusses rather on the theoretical, methodological, and practical problems that arise in international cooperative research involving systematic cross-national comparison. The OCSR experience becomes a fascinating case study, reported in sufficient detail to be provocative and helpful to others who, it is fervently to be hoped, will embark on similar ventures. Exemplifying both of the two principal social psychological methodologies, the sample survey and the small group experiment, the OCSR studies provide an exceptionally relevant case for examination.

I also take great satisfaction in recording the selection by SPSSI Council of Dr. John Harding, of the Department of Child Development and Family Relationships, Cornell University, as the new General Editor of this *Journal.* Dr. Harding brings to his editorship a background of long and intimate association with the *Journal.* He has served loyally on the Executive Editorial Committee, and himself edited the issue on "Intergroup Conflict and Racial Attitudes" (Vol. 8, No. 1, 1952). I can think of no one who better exemplifies the difficult dual commitment to scientific adequacy and social relevance to which this *Journal* is dedicated. The *Journal* is fortunate indeed to come under the guidance of so firm a believer in its objectives and so competent an executor of them. While I will continue as general editor through the next issue, correspondence about future issues should henceforth be directed to Dr. Harding in Ithaca.

M. BREWSTER SMITH
General Editor

Foreword

More than thirty European social scientists and a half dozen American colleagues in Europe on Fulbright grants worked together to produce the cross-national research project described in these pages. Only a few are represented in the authorship of this report or the accounts of the research that are completed. This imbalance will be remedied in other planned publications. Many of the teams have written articles for journals in their countries. Some are doing special analyses that will be worked through cross-nationally for later publication. The cross-national analysis of the survey data, from which an illustrative section was adapted for this Journal, is being completed in the United States this spring and the results of these analyses may be published in a single volume at a later date. Since this is the first public presentation of the survey material, the articles concerned with the survey concentrate on a statement of the research design and a presentation of some of the results of first analyses of this material. A detailed article on the group experiment appears in Volume VII, Number 4 of *Human Relations*. Rather than repeating this material, the two articles stimulated by the experiment treat interpretive and methodological problems which are relevant to the design, execution and understanding of any cross-nationally replicated research.

Because this issue could have been written by any of the senior scientists in the cooperating research teams, we are not presenting the statement about authors that is ordinarily found in this *Journal*. In its place we list the senior research personnel of the fourteen teams composing the Organization for Comparative Social Research, the central coordinating staff of the organization and the Fulbright scholars who acted as consultants and coordinators during the planning and execution of the research program.

The National Research Teams:

	Group Experiment	***Survey***
Belgium:	University of Louvain Laboratorium voor Experimentale Psychologie	University of Louvain Institut de Recherches Economiques et Sociales
	Josef Nuttin, Research Director	*Claire LePlae,* Research Director
	Rafael Kluppels	
	Valère Lamsens	

France:	Centre d'Etudes Sociologiques, Centre National de la Recherche Scientifique, Paris	
	Paul Maucorps, Research Director	*Paul Maucorps,* Research Director
	Jacques van Bockstaele	Institut National de la Statistique et des Etudes Economiques, Paris
		Marcel Brichler
Germany:	Johan Wolfgang Goethe Universität, Frankfurt/Main; Institut für Sozialforschung	Institut für Demoskopie, Allensbach/Bodensee
	Diedrich Osmer, Research Director	*Ulrich Jetter,* Research Director
	Volker von Hagen	
	Werner Mangold	
	Christoph Oehler	
	Inge Rauter	
	Peter Schönbach	
Netherlands:	University of Amsterdam, Psychologisch Laboratorium	Carolus Magnus University, Nijmegen; Sociologisch Research Centrum
	Hubertus Duijker, Research Director	*Antoine Oldendorff,* Research Director
	Mauritz Mulder	*P. O. M. Brackel*
Norway:	University of Oslo; Institute for Social Research	University of Oslo; Institute for Social Research
	Ragnar Rommetveit, Research Director	*Vilhelm Aubert,* Research Director
	Gerdt Henrik Vedeler	*Harald Ofstad*
		Per Tiller
Sweden:	University of Stockholm	University of Stockholm
	Joachim Israel, Research Director	*Edmund Dahlström,* Research Director
	Irene Steg	*Eskil Björklund,* Research Director
United Kingdom:	University College, London	London School of Economics and Political Science
	Cecily de Monchaux, Research Director	*Hilde T. Himmelweit,* Research Director
	Sylvia Shimmin	*R. S. Schenk*

The Central Administration Office:

Erik Rinde, Chairman	Institute for Social Research, Oslo
Hubertus Duijker, Deputy Chairman	University of Amsterdam, Psychologisch Laboratorium
Stein Rokkan, Executive Secretary	University of Oslo, Institute for Social Research

Fulbright scholars associated with the project:

David Krech	University of California	1949-1950
Herbert Hyman	Columbia University	1950-1951
Daniel Katz	University of Michigan	1951-1952
Stanley Schachter	University of Minnesota	1952-1953
Eugene Jacobson	University of Michigan	1952-1953
Burton Fisher	University of Wisconsin	1953-1954

A research enterprise of this degree of complexity could not have been conducted without the aid, support and guidance of numerous institutions and individuals in Europe and the United States. We are particularly grateful for the generous support of the Ford Foundation and for the encouragement of UNESCO and SPSSI. None of the plans developed at the Oslo Seminar for Comparative Social Research could have been carried out without the aid of the Ford Foundation, Behavioral Science Division. We wish to record our gratitude to the director of the division, Dr. Bernard Berelson, and to Dr. Donald G. Marquis, consultant to the division at the time the grant was made. Alva Myrdal and Otto Klineberg of the Social Sciences Department of UNESCO have maintained sympathetic interest in this venture at every stage of its development. SPSSI's efforts to advance international cooperation in the behavioral sciences also played a major role in these developments. The idea of calling an International Seminar to plan comparative research was developed in conversations between Erik Rinde and David Krech, then President-elect of SPSSI.

EUGENE JACOBSON
STANLEY SCHACHTER

Introduction

Daniel Katz and Herbert Hyman

The solution of practical and scientific problems too often has a history of tortuous and circuitous efforts before the logic of a naive, straightforward approach asserts itself. The direct approach carries the threat of actually facing up to the problem and meeting its difficulties. Comparative research in social psychology has suffered from this general unwillingness to face squarely the primary methodological issues. There has been no frontal attack upon substantive problems with a methodology involving similar samples, common research design, common procedures and functionally equivalent measures. Anthropological studies have done much to make us aware of the culture-bound character of our discipline, but our attempts to build systematic knowledge on the basis of comparative and cross-national research have been feeble and fragmentary. We have devoted considerable energy to interpreting qualitative anthropological material but have done little to impose controls in the gathering of the data to insure comparability and adequate quantification. There have been occasional attempts at replication in which one investigator tries to carry through in some fractional manner part of a study carried out in another nation. But again no true comparability of method has been achieved. The samples are generally not comparable and sometimes have no known characteristic, and the procedures and methods vary widely.

The results of comparative studies will always be ambiguous and will comprise an area of speculative interpretation rather than systematic scientific knowledge until we can attack problems with the same essential methodology in different nations and cultures that we apply within an experimental or research program in a single country. New techniques and procedures may have to be devised but the basic experimental logic which calls for controls in the gathering of data, for quantification of variables and for manipulation of variables through experimental or statistical procedures can not be suspended merely because the research site includes more than one country. The major purposes of cross-national research are to test the extent to which principles found in one culture or sub-culture can be generalized to other cultures, i.e., to discover the limiting conditions of our generalizations and to find out the effects of cultural variables on social-psychological processes. At some point then we need to move boldly and directly to set up studies which maximize comparability of research designs and procedures.

The attempt to achieve such comparability is the point of departure of the project reported in the following pages in which social scientists

from seven Northwestern European countries organized themselves into *The Organization for Comparative Social Research* to plan and carry through a seven-country study as a unified piece of research.

The problem which the social scientists from the seven countries selected for investigation was the effect upon group processes of perceived threat to the group. The original interest of most of the members had been broader and had been concerned with the causes of national and international tension. It was soon realized that any thorough exploration of the causes of war with its necessary economic and historical investigation would be too ambitious and too long-term a program for a first project in cross-national research. Accordingly, it was decided not to study the original causes of tension but to accept the existence of national tensions arising from perceived threat as a fact and to investigate the social-psychological effects of these tensions. This formulation of the problem not only made it researchable with respect to national tension but also permitted equivalent types of hypotheses to be tested statistically through nation-wide surveys and experimentally at the small group level.

Thus the basic hypothesis under investigation was that, as perception of threat to the group increases, the central values of the group become more salient and the pressures toward conformity become greater. These effects may be manifest in a narrowing of group norms, in a redefinition of the boundaries of the group, in more actual felt pressure to conform, in a perceptual distortion of the threatening outgroup, etc. This is basically the old position of Felix le Dantec who held that every social unit existed only because there was a common enemy. In fact, one major group of theorists has traced the rise of the state and of all coercive organizations to either the fear of an out-group or the need to exploit the out-group (Gumplowicz, Spencer, Oppenheimer). And as Gordon Allport points out, the resort to building up a perception of threat from the out-group has become a well-known political technique. Hitler would have been greatly handicapped without the presence of the Jewish group he tried to exterminate. McCarthy's fantastic rise to national prominence and power rested largely upon his securing acceptance for his role of attacking traitors during a period of national insecurity. Orwell in *1984* presents an extension of this theme when he describes a world situation in which a dictatorial regime maintains some degree of localized warfare and hence a constant condition of threat and danger to the nation to reinforce the devastating conformity pressures vital to the existence of the dictatorship.

Although it is possible to find ready acceptance for the thesis that external threat makes for group solidarity of an intolerant quality, it is still true that we lack systematic evidence about the extent to which this generalization holds, about the nature of the limiting conditions, and about the processes which bring it about. In World War II the threat of German invasion to France seemed to work in the opposite fashion, namely to increase divisionism within the social structure. It is important, therefore,

for social research to carry out some experimental analysis of the problem which will make explicit the nature of the threat and the nature of the group structure which is likely to be affected. Thus the present cross-national project was envisaged as an attempt to move into this broad problem area with a virtual identity of approach with surveys and group experiments in the seven countries. Though the interest was centered about the Orwell thesis it was realized that the survey material would afford a rich safety factor. It would provide comparable material from seven nations on such intervening variables as personality factors and child rearing practices and on dependent variables such as attitudes toward out-groups and tolerance for non-conformity, and in this way make possible valuable types of secondary analysis.

Organizational Aspects of Cross-National Social Research

H. C. J. Duijker and Stein Rokkan

International cooperation in the planning, conduct and utilization of research in the social sciences has taken on increasing importance during the years since the end of World War II, but very little has yet been done to analyze the methodological and organizational problems of such research cooperation. More and more social scientists have begun to see their theories, their data and their findings in cross-cultural and cross-national perspectives, but few have given much thought to the problems of effective organization for comparative research across cultural and national boundaries.[1]

On the following pages an attempt will be made to clarify some of these problems as they appear after three years' experience in the administration of a program of cross-national social research in Western Europe. It is our hope that these notes will prove of some use to other social scientists planning to engage in similar cross-national ventures, whether in Europe or in other regions of the world.

Cross-National Social Research: Some Distinctions

"Cross-national research" is a relative newcomer in the vocabulary of social science. It is gaining acceptance as a convenient term for comparative inquiries taking nations, national populations, or equivalent sections of national populations as their units of comparison. The term "cross-national" seems to have been coined by analogy with "cross-cultural," a favorite adjective in anthropological discourse since the 'thirties.[2] "Cross-national" was presumably also introduced to reduce some of the ambiguities of the term "international" as used in describing research. This term hides a distinction of some importance in discussing alternative organizational arrange-

---

1 Among the few relevant items in the literature can be mentioned M. Fiske and L. Lowenthal, "Some Problems in the Administration of International Communications Research," *Public Opinion Quarterly, 16* (2), Summer, 1952, 149-59; George P. Murdock, "The Processing of Anthropological Materials," in A. L. Kroeber, ed. *Anthropology Today,* Chicago, University of Chicago Press, 1953, 476-487; John W. M. Whiting, "The Cross-Cultural Method," in G. Lindzey, ed. *Handbook of Social Psychology,* Cambridge, Mass., Addison-Wesley, 1954; Margaret Mead and Rhoda Métraux, eds., *The Study of Culture at a Distance,* Chicago, University of Chicago Press, 1953, 85-101.

2 Cf. George Peter Murdock, "The Cross-cultural Survey," *American Sociological Review, 5* (3), 1940, 361-370; André J. Köbben, "New Ways of Presenting an Old Idea: The Statistical Method in Social Anthropology," *J. Royal Anthropological Institute, 82* (II), 1952, 129-146.

ments: the distinction between *cooperative* research among social scientists or social science centers of different countries—and *comparative* research on different national populations. Cooperative international research need not be comparative: social scientists from different countries could work together in one country on problems of common concern and research centres in different countries could develop joint programs of complementary studies without engaging in any comparisons between national populations. And, *vice versa,* comparative research need not involve between-nation co-operation—especially if the data required were already available for the countries to be covered and were only waiting to be analyzed. Some terminological differentiation seemed to be required to cope with this distinction. "International research" tended to become associated with the cooperative aspect, and "cross-national research" came to take its place as a term for between-nation comparative studies. Usage is still very unstable, but in the present context we feel that "cross-national research" can serve as a convenient shorthand term for the kind of social science activity we wish to discuss: *research undertaken for comparative purposes on the same categories of data across several different national populations or equivalent sections of different national populations.*

This is clearly a very wide concept: it covers an immense range of possible operations, structures, purposes. Cross-national research can take the most different forms, can involve a great variety of activities, and can require highly divergent types of organizational arrangements. An adequate typology of research set-ups is not easily devised: projects can differ in a great number of dimensions.

For the purposes of the present discussion of organizational problems it would seem important to distinguish at a first level between three principal types of comparative cross-national research:

I. *"Documentary" studies:* comparative analyses of characteristics and relationships in already existing records and materials.

II. *"Current statistics" studies:* comparative analyses made possible through increased standardization of data collection and classification procedures in regularly operating statistical agencies and other organizations for the registration of social facts.

III. *"Field and laboratory" studies:* analyses of data specifically collected and classified for the comparative purposes in mind, whether through direct observation, interviewing, test administration, field experiments, or laboratory experiments.

Organizational requirements must necessarily differ between these three types of cross-national research: differences will be even more striking when programmes involving the possible combinations of these three types are considered. *"Documentary"* studies do not require any direct contact with the national populations to be analyzed: once the records are accessible, there is no problem of data collection—the task is to classify and analyze the data as they are historically given. *"Current statistics"* studies do require contacts with the national populations, but through established stock-taking and record-keeping organizations regularly collecting the data for their own national purposes—the task in this case is to assess the possibilities and

limitations of currently collected data for purposes of comparative research and to improve the bases for such analyses through increasing international standardization of procedure. *"Field and laboratory"* studies, finally, require direct contacts with the national populations in collecting data otherwise not obtainable—the task is to develop data collection procedures directly adapted to the objectives of comparative analysis, to gather data for the analysis in mind, not to analyze whatever data are already available or are currently collected for other purposes.

Examples of these three types of cross-national research will illustrate how these differences in data collection-data analysis relations are reflected in the cooperative arrangements required.

Type I. Cross-national studies of Type I of course abound in the literature, but the majority of such studies leave much to be desired from a social science point of view, although they certainly constitute important sources for hypotheses and research ideas. Some of the cross-national studies of this "documentary" type have been one-man projects and have not posed any organizational problems beyond those of efficient utilization of existing library and archive facilities. It is clear, though, that all extensive studies of documentary materials will sooner or later require some team work and institution-building and consequently pose problems of research organization. In the field of cultural anthropology, an interesting example of such cooperation is of course the Human Relations Area Files.[3] In the field of political sociology, an important example of a cross-national research programme of the "documentary" type is afforded by the RADIR studies at the Hoover Institute and Library on War, Revolution and Peace at Stanford University.[4] This attempt to investigate on a comparative basis the characteristics of the changes that have occurred in "the *composition* and *vocabulary* of the wielders of power" in a series of selected countries involved extensive use of available records of political developments in each of the countries: biographical documents providing data on the careers of members of old and new elites, communication materials, primarily newspaper editorials and political speeches, reflecting changes in argumentation, modes of appeal, uses of keywords. Most of these materials were available at the Hoover Library or in other libraries and archives in the United States and could be examined for purposes of coding and analysis by researchers operating in one organizational unit. Problems of decision-making, administration and supervision, therefore, differed very little from those of social science institutions operating on a national level. This, however, would not

---

[3] G. P. Murdock *et al, Outline of Cultural Materials,* 3rd ed. New Haven, Human Relations Area Files, Inc., 1950. Cf. the discussion of the Dutch sociologist Steinmetz's earlier attempts in this direction in Köbben, *op. cit.*

[4] *Hoover Institute Studies,* Palo Alto, Stanford University Press, 1951-53, particularly Ser. A, No. 1, Harold D. Lasswell, *The World Revolution of our Time,* 1951; Ser. B, No. 1, Harold D. Lasswell, Daniel Lerner and C. E. Rothwell, *The Comparative Study of Elites,* 1952; Ser. C, No. 1, Harold D. Lasswell, Daniel Lerner and Ithiel deSola Pool, *The Comparative Study of Symbols,* 1952.

necessarily apply in all cross-national research arrangements within our Type I: considerations of methodology might in some cases make it more appropriate to have national teams in charge of content analyses of their respective documentary data within some comparative framework. In such cases problems of between-nation relations and between-team interaction would arise and problems of cross-national comparability of research procedures would become more acute.

Type II. Organizational problems are of a different sort for cross-national studies of our Type II. An enormous number of demographic, economic, political and sociological inquiries come under this head: they take the form of between-country comparisons of data provided through the regular bookkeeping operations of public and private agencies of different kinds. Decades of interaction and negotiation among census bureaux, national statistical agencies and other public and private stock-taking organizations have resulted in increasing international comparability for a great number of data categories: population, health, production, occupation, income, consumption, standard of living, trade, education, etc. The United Nations and the Specialized Agencies have made important efforts to accelerate this process of international standardization and codification of research procedures in current statistics, and have issued a number of manuals and classification schemes for adoption by national agencies.[5] Cross-national research has an important role to play in this process. Comparative inquiries call attention to the need for standardization, indicate possible equivalences, and can be of assistance in engineering coordination schemes acceptable to the parties concerned. Such research will take different organizational forms according to purposes and conditions: it will often be closely related to the planning and administration of general cooperative arrangements between national institutions.

Type III. Cross-national studies of our Type III are much fewer in number, but there is reason to believe that they will grow in frequency over the coming years with the further development of international cooperation in the social and behavioral sciences. Comparative sample inquiries, opinion polls, attitude surveys, ecological researches and community studies, organizational investigations, group experiments, and psychological testing programmes across national boundaries are likely to take on increasing importance over the next decade. Cross-national research of this kind requires direct contact with the national populations through field work, test administration or laboratory experiments, and will depend on organizational arrangements very different from those involved in "documentary" and "current statistics" studies.

Most Type III studies must of necessity be limited in time and scope

[5] W. R. Leonard, *Proc. Internat. Statist. Conf. 1947.* United Nations, 1949, vol. II, 201-13. *Problems in the Collection and Comparability of International Statistics.* New York, Milbank Memorial Fund, 1949. Also: United Nations. *Statist. Papers.* Ser.M. No. 16. March 1953.

and cannot be based on continuous or regular-interval operations. In the field of public opinion polling and for certain types of psychological test administration there are, however, tendencies in this direction of continuity and regularity in the data gathering operations for a few countries: in this sense, they would be on the borderline between our Types II and III.[6] The prospects for regularization and routinization of cross-national research for the other methodologies are still distant. A great number of difficult problems of procedure, coordination and comparison remain to be clarified and solved, and much exploration and experimentation with different types of research arrangements must take place before cross-national data collection can be regularized within stable institutional frameworks.

Two Dimensions of Cross-National Research

Whatever the regularity of operation, the problems of *time and timing* remain of essential importance in cross-national research. Most cross-national studies of our Type III are based on data gathering operations *succeeding each other* at different points of time: a research design, a field instrument, a test, a set of questions have been used with some success in one country, are subsequently adapted for use in one or several other countries and thus provide the basis for some comparative analysis of findings. Less frequent than these *successive* data gathering operations are *concurrent* ones for cross-national comparison purposes: the same research design and instrument are used in several countries over roughly the same period of time. This distinction between successive and concurrent operations has important implications for research organization. Successive data gathering operations have in most cases come about without much formal cooperation: the normal condition has simply been the availability of the documents and procedural records of the previous study—in some cases roughly comparable data have been gathered even without this minimum of interaction—through the "accidental" coincidence of similar research concerns. Concurrent data gathering operations, on the other hand, have normally required extensive organizational arrangements for interaction and coordination at all stages of the research. The rationale for the organization of concurrent cross-national data gathering operations is two-fold: it normally makes for stricter comparability, and it makes it possible to keep the time factor constant. The importance of the time factor of course varies enormously with the kind of variables and relationships under study, but there can be little doubt that it is essential to keep it controlled in studies dealing with opinions, attitudes and ideologies deriving some of their content from current social and political developments.

This "successive-concurrent" distinction cuts across another dichotomy of equal importance to the discussion of organizational aspects of cross-

[6] An interesting example of cross-national analysis based on opinion data as well as "current statistics" data collected for national purposes is Natalie Rogoff, "Social stratification in France and the United States," *Amer. J. Sociol., 48* (4), Jan. 1953, 347-357.

national research: the distinction between *"repetitive"* and *"joint development"* research. This distinction concerns the mode of decision-making determining the design and procedure of the comparative country studies. In *"repetitive"* research the design, the instructions and the instruments have been developed and perhaps already used by one initiator, a researcher, a team or an institution, in one country and are repeated in as equivalent a form as possible in studies in one or several other countries. In *"joint development"* research the design and the procedures of the cross-national study are decided on in common by an internationally constituted body of researchers or representatives of research institutions.

A Further Typology

Combining these two dichotomies we get four major sub-types within our Type III. The "repetitive-successive" type of cross-national research is perhaps the most frequent: numerous studies within differential psychology fall in this category. The "joint development-successive" type is much less frequent and can only take place under the sponsorship of a stable organization operating on a long-term basis. An example of such research might be the programme of comparative research on social stratification and social mobility initiated by the International Sociological Association in 1951[7]: this took its point of departure in the British mobility study conducted under the direction of David Glass at the London School of Economics,[8] involved a series of discussions among experts from countries interested in undertaking similar researches, and led to a series of successive national projects incorporating some of the procedures and the instruments used in the British study.[9] The "repetitive-concurrent" type of cross-national research is again more frequent. International public opinion polls are largely of this type: a central firm or agency specifies the purposes of inquiry and the questions to be asked and commissions national polling or marketing organizations to carry out the interviews as specified.[10] The UNESCO nine-country public opinion poll directed by Hadley Cantril would also seem to

---

[7] Cf. Erik Rinde and Stein Rokkan, eds., *First International Working Conference on Social Stratification and Social Mobility: Preliminary Papers and Proposals,* Oslo, International Sociological Association, 1951, 154 pp. Further papers are assembled in *Transactions of the Second World Congress of Sociology,* Vol. II, London, International Sociological Association, forthcoming 1954. A Third International Working Conference is to take place in December, 1954.

[8] D. V. Glass, ed., *Social Mobility in Britain,* London, Routledge and Kegan Paul, 1954, 412 pp.

[9] Such studies have been completed or are in progress in Japan (Kunio Odaka, University of Tokyo), the Netherlands (F. van Heek, State University of Leiden) and Denmark (Kaare Svalastoga, University of Copenhagen).

[10] Cf. Laszlo Radvanyi, "Problems of International Opinion Surveys," *Int. J. Opin. Attit. Res., 1,* 1947, 30-51; D. Wallace *et al.,* "Experience in the 'Time' International Survey: A Symposium," *Publ. Opin. Quart., 12* (4), 1948-49, 709-721; "Proceedings of the World Association for Public Opinion Research," *Int. J. Opin. Attit. Res., 3,* 1948, 309-335, and *4,* 1950, 259-309; *Int. Soc. Sci. Bull., 5* (3), 1953, 520-554, especially "Problems of International Polls," pp. 534-536, "Techniques in International Polling," pp. 542-546. WAPOR is currently preparing a "manual" of international polling under a contract with Unesco.

belong to this type, although there were also some "joint development" features in this project.[11]

Very few of the cross-national research programmes thus far recorded can be said to belong to the last of the four sub-types: the "joint development-concurrent." Research programmes of this kind require a more elaborate organizational apparatus and involve intensive interaction between the participating researchers and teams at every stage: from the early planning phases, through the design and instrument construction phases, the data gathering and the data processing, to the final stages of comparative analysis, interpretation and report preparation. UNESCO's programme of rural and urban community studies in four countries is one example of this type of research. This involved teams from four countries as widely dispersed as Australia, France, India and Sweden in the planning of comparative community researches.[12] In the actual implementation of these research plans, however, the enormous distances between the countries and the scarcity of funds available made it very difficult to keep between-team interaction at the level that was required to ensure comparability: the result was a series of very interesting national studies and no comparative report. Another example of research of the "joint development-concurrent" type is the programme of the *Organization for Comparative Social Research*. Since this group paid particular attention to problems of communication, cooperation and coordination, it may be worth while to consider its orgaanizational features in more detail.

The OCSR: Notes Toward a Case Study

The Organization for Comparative Social Research grew out of a series of exchanges, discussions and agreements among Western European social scientists convened for successive sessions of an International Seminar set up by the Institute for Social Research in Oslo in 1951. The aims of this International Seminar were twofold: 1) to ensure detailed discussion and effective agreement on a plan for the development and conduct of a set of comparative inquiries to be carried out by national teams within some cooperative framework; and 2) to provide specialized training in behavioral science methodologies through such discussion and planning and to ensure

---

11 W. Buchanan and H. Cantril, *How Nations See Each Other,* Urbana, University of Illinois Press, 1953.
Another example of an international study of this type is Gordon Allport's ongoing study of the attitudes of youth as expressed in essays on "The World in the year 2000."

12 Cf. Adam Curle, "A Conference on the Methodology of Social Surveys," *Int. Soc. Sci. Bull. 3* (3), 1951, 629-634. Results of the Australian inquiries are presented in O. A. Oeser, F. E. Emory and S. B. Hammond, *Social Structure and Personality* I-II, London, Routledge and Kegan Paul, 1954. The Swedish studies are reported in Edmund Dahlström, *Pinehill: a Community Study in a Big City,* Uppsala, Sociol. Inst., 1951, 185 pp. mimeo. and in Bengt G. Rundblad, *Forestville: a Study of Rural Social Change,* Uppsala, Sociol. Inst., 1951, 165 pp. mimeo. Further results of the Swedish inquiries are reported in Eskil Björklund and Joachim Israel, *The Authoritarian Ideology of Upbringing,* Uppsala, Sociol. Inst., 1951, 387 pp. (mimeo.).

effective research training on the national level through participation in the actual inquiries to be carried out. The OCSR was a direct outcome of the planning made possible through the Seminar: the establishment of the OCSR became a necessity when financial resources were suddenly found available for the research programme that had been discussed in the Seminar—sometimes with very little concern for the possibility that it might actually have to be carried out. The structure and the functions of the OCSR were largely determined by the decisions that were taken in launching the Seminar and by the interaction processes that took place during the Seminar sessions.[13]

Initial decisions. Two of the initial planning decisions proved of essential importance in determining not only the organizational pattern but even more the actual research product itself. It was decided that the design and development of the research programme should be the *joint responsibility* of the Seminar members—there would be no imposition of a design already established or a set of procedures already used, the only limitation being that the programme to be decided on would be focussed on problems and relationships of importance in the general area of intergroup conflicts and international tensions. It was further decided that the principal Seminar participant to be recruited from each of the Western European countries to be covered should be a younger social scientist trained in empirical research and specialized in one or more of the current methodologies of behavioral science: there was thus *no strict limitation to one methodology or one type of academic background* but on the contrary explicit emphasis on the interdisciplinary character of the Seminar and on the need to involve both sociologists and social psychologists in the venture.

These two sets of decisions were clearly interrelated and they combined to make the actual process of research planning and research implementation a complex and cumbersome one. In a "joint development" programme of this kind the design and the research procedures are functions of the personnel enlisted: in a "repetitive" programme the opposite is the case—the personnel is a function of the design and the procedures already decided on. In the case of the OCSR, there was not only joint development but also a characteristic heterogeneity of membership—in terms of academic background, theoretical orientation as well as research skill and experience. This heterogeneity was immediately reflected in the interaction processes within the seminar and led to disagreements which could only be solved through compromises and structural divisions within the membership framework once established.

Recruitment. This heterogeneity of membership was not exclusively an outcome of the original decision to emphasize interdisciplinarity; it also reflected very strikingly the unevenness of social science developments and training facilities in the countries of western Europe. Actually, the seminar

[13] For an account of the Seminar, see Stein Rokkan, *Institute for Social Research, Report of Activities 1950-52,* Oslo, Institute for Social Research, 1952, pp. 29-40. Mimeographed *Proceedings* of the Seminar were issued in five installments edited by Björn Christiansen and Stein Rokkan.

found part of its justification in the need to improve training opportunities for younger social scientists through concerted international action. Participation in the planning, designing and conduct of comparative inquiries in a number of countries would not only provide opportunities for the training of local researchers, field workers and students but also afford an excellent means for the testing of research standards and the improvement of conditions and facilities. There was thus a clear relationship between the comparative research plans, the stress on the training features and the recruitment policy for the Seminar. In the circumstances, this combination of emphases seemed the most fruitful, but in future cross-national research programmes it should be possible to delimit the problem area much more rigorously and to adapt the recruitment more directly to the requirements of research in that area.

The original plan was to invite one social scientist from each of ten Western European countries to take part in the Seminar. Recruitment letters were sent to professors and research directors in the fields of psychology and sociology in all the countries concerned and requests were made for advice on suitable candidates for participation. As criteria to be used in the selection of the candidates, it was proposed 1) that he or she be interested in the application of sociological or psychological research to the problems of international tensions; 2) that he or she have reached an advanced post-graduate level of training without having advanced so far as not to be likely to profit from the training features of the Conference; 3) that his or her area of specialization be within sociology or social psychology; 4) that his or her orientation be in the direction of empirical research methods; and finally 5) that he or she have an adequate knowledge of the English language.

As could be expected, the candidates proposed by the professors and research directors consulted varied strikingly in academic background, research experience and general orientation. It became clear that it would be very difficult to find suitable candidates in all the ten countries originally envisaged and that the countries to be covered would have to be restricted to seven or eight. To provide a better basis for the final selection of Seminar members, all the proposed candidates were requested to give detailed information about their primary research interests, their research experience and the research resources and facilities they would have access to in their country if a comparative inquiry were to be launched. Research experience was to be specified for a number of social science techniques: design of samples, interviewing and field work, supervision and training of interviewers, coding and content analysis, statistical analysis of data, institutional analysis on the basis of records, mass media, statistical series, participant observation. Research resources and facilities were to be specified in terms of institutional framework, availability of research personnel, senior advisers, trained interviewers and field workers, clerical assistants and student volunteers, statistical machinery, population listings and other materials for

sampling, possibilities of local research support. Even on the basis of the detailed data received in this way, it was far from easy to make a definite selection of Seminar members, and it was clear that it would be practically impossible to establish a group of really "articulating minds." Not only would it be difficult to find people with compatible research interests: it seemed even more difficult to find candidates who had the kind of research experience required and at the same time would have access to the institutional facilities that would be required if comparative inquiries were to be launched. These difficulties clearly reflected the unevenness of developments in the social and behavioral sciences in the Western European countries.

Early interaction. After considerable discussion, it was agreed to invite ten candidates from eight countries to a preliminary conference in Oslo in May, 1951. Representatives from seven of these countries were all found to be motivated by genuine interest in this cross-national research venture and came to constitute the nuclear group in the organization to be established a year later. It was certainly a mixed group and prospects for reaching effective agreement on cooperative research operations were not the best. Backgrounds and skills differed strikingly: there was a naval officer turned sociometrician, a physicist turned public opinion expert, a philosopher also working in social psychology, a lawyer and sociologist, etc. These differences did not prevent them from sharing some faith in the potentialities of cross-national research and quite a bit of enthusiasm for the idea of carrying out comparative inquiries within a cooperative international framework. Without assuming some such common motivation it is at any rate difficult to explain why the group did not fall apart when major divergencies of research orientation appeared during the first days of discussion and interaction.

Major divergencies in research orientation might have been predicted from the data assembled from the candidates before the Seminar was set up. The plans for the Seminar had a strong interdisciplinary emphasis. There was no explicit restriction to one methodology, but the questionnaires used in eliciting data from the candidates gave clear priority to sample surveys and interview inquiries: this will appear from the specifications already given in the account of selection procedures. In the end, however, it proved impossible to limit the Seminar to researchers primarily interested in this methodology: this would have meant excluding some of the most qualified candidates and would indeed have restricted the scope and appeal of the Seminar in a way that had not been intended in the original plans. The result was inevitable—all the Seminar participants were interested in some way or other in the use of sample surveys and interviews in cross-national research, but only about half of them were prepared to give priority to this methodology: the others evidenced stronger interest in experimental studies with groups and advocated the need for a combined approach in cross-national research. This difference in methodological orientation gave rise to very lengthy discussions not only at the preliminary conference but

also at the main conference, a four-week session, in the autumn of 1951. A number of combinations of research designs and research operations were discussed in much detail and at some points there was real danger that the group split apart. At the end, it nevertheless proved possible to reach tentative agreement on a joint design for two parallel sets of cross-national studies to be conducted within one cooperative programme, the one a set of group experiments, the other a set of interview inquiries. This meant establishing two research teams for each of the seven countries and it meant adjusting the Seminar membership to two representatives from each country, one to be in charge of the group experiment, the other in charge of the interview surveys. The additional members recruited to fit this pattern did not have the same amount of influence on the over-all design as the original members: for them the programme had some of the characteristics of what we have called the "repetitive" type of cross-national research since their recruitment was a function of the over-all design agreed on in the original group. The details of the design and the specifications of the procedures and instruments were, however, very far from completed at the time of their recruitment: within the framework already set they all had ample opportunity to influence decision-making at all stages of the operation—through correspondence, through the conferences held, and through pretests and trial runs in their own countries.

Decision-making. The third session of the Seminar was held for ten days in April, 1952. The programme had come close to the point of action. The teams had been set up in each of the countries, some pretesting of instruments had been done, concrete proposals for the first phase of operations had been worked out, the fourteen team leaders were ready to discuss them, and last but not least, the financial basis for the cross-national operations had been secured—the session was opened with the announcement of a Ford Foundation grant toward the completion of the first eighteen months of the three-year programme planned and agreed on in the Seminar. This meant hard work for the Seminar members and it meant interacting under real pressure. For each of the parallel sets of cross-national studies agreements had to be reached on every step in the research process, time tables had to be set up for the operations, measures for the coordination of team activities had to be decided on. Again the difficulties of international decision-making were in evidence. At the first stage of Seminar interaction, disagreements had been focussed on issues of basic theory and general methodological orientation. At this second level of decision-making, disagreements had immediate operational referents: the over-all agreement on the structure of the research programme was respected, but serious difficulties arose in the specification of design elements, procedures, instructions and instruments. In this process of decision-making among team leaders from different countries, many forces were at work. The forces that motivated the individual team leaders in their interaction could easily conflict: their personalities, their career concerns, their institutional roles in their national contexts, their status in relation to their national co-workers. The process of

consensus was necessarily complex and cumbersome: not only was there occasional opposition of interests, but there were also important cultural and linguistic barriers that prevented efficient cooperation.

Communication problems. Language difficulties played their part. It had been decided to use English as a working language. All the Seminar members were able to speak and understand this language, but there were very definite differences in their ability to use the language in debate, to argue their cases and to explain their reasoning in an efficient way. Some of the team leaders were clearly handicapped by their language, they felt that they did not succeed in making their points as they should and they felt frustrated because they were not able to influence decisions the way they would have wanted. It might perhaps have helped to have had interpreters on some occasions, but normally this would have tended to make interaction even more cumbersome. On the whole, cross-national research requires significantly more communication among the participants than national research to get under way. Every point has to be gone over and argued in detail since so very little is taken for granted or understood by implication. Research cooperation can only become efficient after much and often redundant interaction ensuring full understanding of agreements.

In the case of the OCSR, we have seen that it took three conferences to reach the operational stage. Yet another conference was required to discuss the final instructions and instruments before the actual field work could start. To the outsider this might seem an excessive number of conferences, but in actual fact there ought to have been even more opportunities for interaction between the teams to ensure comparability in the research processes. Teams were normally represented "at the top" at these sessions: at the three first conferences there were only senior researchers, but at the fourth a few of the research associates were also able to attend. To ensure a higher degree of comparability, it would probably have been very useful to have involved the experimenters, the observers, the field supervisors, and perhaps even the principal coders in some joint conference to discuss concrete problems of data collection and categorization, but this was clearly impossible within the financial limitations imposed. Conferences were essential because they allowed much more intensive interaction and intercommunication than could ever be achieved by mailed exchanges. Team leaders got to know each other better with each conference and were able to cooperate much more effectively as they became acquainted with each other's modes of expression and understood each other's orientations and frames of reference. To make a success of international research such interaction processes are essential: the OCSR experience was certainly that the less formal exchanges in the meetings and during on-the-spot consultations were much more effective in ensuring common understanding and comparable operations than the more formal communications to and from teams through the mails.

"Common" vs. "nation-specific" features. Given all these difficulties,

how was it at all possible to reach agreement on a common set of operations for all the countries to be covered? The genuine spirit of cooperation evidenced by the Seminar members of course meant very much, but the overriding factor was simply the need to achieve comparability between the national research operations. Agreement on common procedures, instructions and instruments was the *condition sine qua non* of the entire enterprise. There were at times violent disagreements in the group, but there were at the same time very distinct and very strong pressures for conformity: the group goals were set and agreed on and implied definite restraints on dissent. The group had defined itself by its purposes and could only tolerate disagreements within the framework thus set; to go beyond it simply meant leaving the group. This still left a broad margin for conflicts on procedures. At no point was any proposal set forth to abandon comparability and simply finance a series of national studies, but there was much discussion of different possibilities of combining comparative research interests with national interests. This gave rise to the very popular distinction between a "common core" of strictly comparable elements in the research processes, and a series of "nation-specific" elements in these processes. In actual fact, the OCSR studies were very heavily concentrated on "common core" elements, but the very possibility of adding "nation-specific" features to each team study served as a kind of safety valve in the discussion and relieved some of the hardships of rigorous conformity to an international agreement.

This need to strike a balance somewhere between the requirements for a comparative "common core" and the demands for "nation-specific" research elements poses a crucial problem in the organization of cross-national social research. Unless the cross-national arrangements are to be entirely of the "repetitive" type and only involve the commissioning of standardized replications in one country after another, cooperative processes and structures have to be devised that will ensure a safe minimum of strictly comparable elements and at the same time allow adequate scope for local autonomy.

This is not solely a problem of efficient organization. There are also important methodological considerations that argue the need for an emphasis on nation-specific inquiries within a framework of cooperative research. Any common agreement on procedures and instruments will have to be adapted to the national conditions. There is a need for translation not only into another language, but even into another culture or another social structure or political context. To ensure a minimum basis for comparisons, equivalences have to be tested and assessed, but data and interpretations have also to be taken into account that will throw light on the unique elements in the historically given settings. In the OCSR attitude survey there was a definite need to adapt some of the interview questions to the national situation. In Germany, for instance, some of the questions relating to national conflict and defense efforts had to be considerably modified since the literal translations would not make sense in the German situation. The broader

problem was involved of the relationship—organizational as well as methodological—between the statistical approaches to comparative research and the historical and anthropological approaches to the description of different national and cultural situations. Ideally, there should be a combination of approaches: analysis of historical records, studies of mass media, participant observation, attitude surveys, perhaps even experiments, within the same framework. But this is still a distant goal. Meanwhile, it is important to experiment with alternative arrangements on a smaller scale: two or three countries to be covered and two or three methodologies to be combined.[14]

Organizational structure. The long series of agreements that were finally reached by the Seminar members on the concrete operational measures to be taken in each country had definite administrative implications: an Organization of the fourteen teams had to be formally established and its Statutes agreed on, an Executive Committee had to be elected, coordinators for each of the parallel sets of studies had to be appointed, and a Central Office had to be set up to take care of over-all coordination, communication, conference planning and finance. The Organization set up was essentially a temporary federation of research teams in seven countries, united for the common task of designing and implementing comparative inquiries. An Executive Committee was needed to take responsibility for the administration of the program between conferences of the full membership. The two coordinators, one for the attitude surveys, the other for the group experiments, had crucial roles to play in the process. They were to take the lead in developing the details of the design and in engineering the final agreements on procedures and they were to ensure maximum conformity to the agreements through regular on-the-spot consultations with the national teams. Without such travelling coordinators it would have been practically impossible to achieve the standard of comparability that was maintained throughout the studies and to obtain reliable records of necessary or accidental deviations from or adjustments of the common agreements. The coordinators had their headquarters and their staffs of helpers at different institutions to ensure some division of labor and to reduce some of the bureaucratic centralization that might easily develop when so much had to be communicated to so many within such limited periods of time. The group experiment coordinator had his headquarters at the Psychological Laboratory at the University of Amsterdam and had at his disposal the necessary staff to ensure the production and distribution of all the relevant research documents to the teams. The attitude survey coordinator had his headquarters at the OCSR Central Office established as a division of the Institute for Social Research in Oslo. This Central Office served as the administrative organ of the Organization and carried the heaviest burden of responsibility for the progress

[14] Cf. Paul F. Lazarsfeld, "The Prognosis for International Communications Research," *Publ. Opin. Qu., 16* (4), 1952-53, 481-490; Edw. Berrol and Olive Holmes, "Survey and Area Approaches to International Communications Research," *Ibid.* pp. 567-578. A good example of such cooperation between an attitude tester and an area specialist is the Japanese study reported by Jean Stoetzel, *Jeunesse sans chrysanthème ni sabre,* Paris, Unesco-Plon, 1954.

of the research programme. Its functions covered the entire range of operations: from the negotiation of team contracts and the management of the Organization's finances through the production and processing of a long series of communications to teams, to the handling of public relations, the organization of comparative analysis work and the preparation of progress reports, preliminary papers and final publications.

The Central office was a separate unit: there was no overlap in personnel with any of the national teams, not even with the Norwegian ones at the Oslo Institute. Administratively and organizationally this was a definite advantage: it was essential to ensure that all national teams should consider themselves on the same footing in the decision-making process. The work of the coordinators was similarly facilitated by this differentiation of functions. The coordinators were recruited from outside and could assume the roles of neutral and non-partisan experts in the actual or potential conflicts between national teams over designs and procedures: they were not to represent one particular research team—their job was to bring about effective agreements on sets of feasible operations and to ensure optimal conformity to these agreements in the actual inquiries. These were difficult and often harassing tasks—if the national findings are at all worth comparing, this is not least due to the scientific skill, the social insight and the diplomatic perseverance of the two coordinators.

Problems of data utilization. Once the agreements on procedures and coordination measures were reached in the OCSR, the national data gathering operations went largely as planned. There were all kinds of difficulties in the field work and the laboratory experiments, but on the whole the data gathering went quite well and was completed within the time limits set. The interview data were coded and punched on IBM cards by the respective national teams according to a joint scheme agreed on and the group experiment data were tabulated in accordance with agreements reached in conferences between team leaders. But what about the next stage? How was the analysis and write-up work to be organized? How were assignments, responsibilities and credits to be distributed? This is another set of problems to be faced by any organization undertaking cross-national social research. These problems are of crucial importance in the development of any such programme: the suggested solutions to these problems of data utilization will necessarily affect the strength of motivation and the intensity of involvement of every one of the researchers engaged in the programme.

In the case of the OCSR group experiments it was decided that a comparative report should be prepared by the coordinator and his assistant as soon as possible after the completion of the national data tabulations and that this central report was to be published before any national reports were to be prepared by the individual teams. This was not an easy decision to take since it meant delaying the publication of national reports and thus affected the professional career prospects of some of the team leaders, but it was found essential to keep the scientific market intact for the joint presentation of the design, the operations and the comparative findings.

In the case of the OCSR attitude surveys, several solutions were considered. At one time it was suggested that all the participating teams receive duplicates of the full seven-country set of IBM cards and be allowed to undertake such comparative analyses as would interest each of them. This, however, was found to be impracticable: not only would there be no guarantee that one central comparative report could be prepared, there was also the definite possibility of contradictory and conflicting analyses and interpretations of the same data getting published without adequate inter-team consultation. Some measure of centralization of the data processing and analysis was therefore found to be indispensable. All the teams were accordingly asked to send duplicates of their national IBM cards to a central processing unit. This unit worked under the direct instructions of the Central Office staff and produced the comparative tables of response distributions as well as a series of cross-tabulations ordered by the analysts in consultation with the national teams.

This procedure differed radically from the one used by international opinion polls such as the UNESCO nine-country study directed by Hadley Cantril. In the UNESCO study, all the tabulation work was done according to a set scheme by each of the national polling agencies—with the results that those in charge of comparative analysis were limited to the tabulations already decided on and had no opportunity to explore additional and more complex relationships in the data in the course of their work.[15] In the OCSR study, the central processing of all the national data allowed a wider scope for the testing of alternative analysis procedures. For purposes of strict methodological control of coding operations in the different countries, it was also considered useful to have all interview protocols and categorization records assembled in central files, but this proved financially as well as organizationally impossible to arrange. All the comparative analysis of the interview materials was undertaken by the Central Office staff, but the national teams were kept regularly informed of developments, were sent preliminary reports and tabulations and were requested to submit comments and questions toward the interpretation of the relationships in the data. Much encouragement was also given to the planning and execution of nation-specific analyses by each of the local teams and a number of independent reports on the national interview data were undertaken and seem likely to supplement in a useful way the central comparative reports to be issued.

The experiences gained in organizing the analysis and preparing the central publications within the OCSR program seem to suggest that a satisfactory solution of this problem of data utilization can only be found within an on-going federation of research institutions cooperating over a longer period of time. In the case of the OCSR, financial and organizational limitations did not make it possible to devise adequate arrangements for the distribution of assignments for comparative analysis and interpretation among a number of the data gathering teams. Within an on-going federa-

[15] Cf. Buchanan & Cantril, *op. cit.* pp. 122-124.

tion of research institutions, it might be possible to devise some kind of rotation of roles and thus ensure stronger motivation for active involvement in comparative research. In one phase a team at institute A could take charge of design and analysis and get teams at B and C to collect specified data in their respective countries, in the next phase a team at B could take the central role and get teams at A and C to collect the data, and so on. Such rotation of "initiator" and "service" roles between institutions in different countries may well be fitted into long-term research programs and combine "common core" comparative inquiries with "nation-specific" studies. Such a system of interdependent relationships between research initiators and data gatherers might in the long run be more fruitful and less harassing than cross-national research ventures of the "joint development type" involving as they often may a series of compromises and a division of labor that in the end may not satisfy any of the parties. The field is wide open for discussion, exploration and experimentation: no satisfactory solutions have as yet been found, but some experiments have been made that do not seem worth repeating.

* * *

These notes on organizational problems in cross-national field and laboratory studies of human social behavior may have seemed unduly pessimistic. Exclusive emphasis has been given to the difficulties that must be faced, the sacrifices that have to be made, the risks that have to be taken in any attempt to obtain reliable data on social behavior for purposes of cross-national comparison. Our paper has taken it for granted that such difficulties are worth facing. We have assumed that the gains to be derived will match the efforts that are required—at least in the long run. We shall not try to justify this assumption in this paper. Some of the arguments were marshalled in the planning memoranda that started off our venture,[16] and a number of concrete indications of the potentialities of cross-national research will be found in the research reports coming out of our enterprise. Much work remains to be done before cross-national research can become a safe proposition. All we can say to our fellow researchers is: don't repeat our mistakes—have another try!

[16] Cf. Herbert Hyman, "The Conditions Warranting Cross-National Research," in Björn Christiansen, Herbert Hyman and Ragnar Rommetveit, *Cross-National Social Research,* Oslo, Institute for Social Research, May 1951, pp. 3-8 (mimeo).

A Comparative Study of Teachers' Attitudes to International Problems and Policies: Preliminary Review of Relationships in Interview Data from Seven Western European Countries[1]

Vilhelm Aubert, Burton R. Fisher, Stein Rokkan

Introduction

We shall present, in rather condensed fashion, a report on some of the early findings in the analysis of data collected through interviews conducted in 1953 with teachers in seven Western European countries. The countries are Belgium, England and Wales, France, the Netherlands, Norway, Sweden and Western Germany.

Final agreement on the survey project was reached at a conference in November, 1952. The seven national teams were to carry out intensive interviews with stratified cluster samples of 300 primary and 100 secondary school teachers in each of the respective countries during the first months of 1953; the teams were to use essentially the same sets of field instructions and interview schedules in each country, were to have all national research instruments checked centrally for problems of translation and equivalent adaptation before actual use, and were to take special precautions to ensure a high level of comparability in all phases of the inquiries through regular correspondence and consultation with the research coordinator on his visits to the participating centers.[2]

The decision to concentrate the studies on teachers in primary and secondary schools was prompted by several reasons. First, the availability of

[1] The investigations reported on in this paper were planned and conducted by seven teams of social scientists cooperating within the framework of the *Organization for Comparative Social Research:* details on the teams are given in the Introduction to the present issue. The collaboration of all the seven teams in the comparative analysis of the data and the preparation of this preliminary report is gratefully acknowledged. Special thanks for their cooperation at various stages of the research are due not only to the coordinators, Professor Katz and Dr. Jacobson, but also to Dr. Sverre Lysgaard, Professor Harald Ofstad and Dr. Ragnar Rommetveit of the Institute for Social Research in Oslo. The authors are also very much indebted to Drs. P. Brackel of the Sociological Research Centre at Nijmegen and Miss Grethe Haldorsen of the Institute for Social Research in Oslo for their work with the processing of the coded data.

[2] See the article by Eugene Jacobson on methods, this issue.

reliable rosters of all teachers or school classes in all countries would simplify sampling procedures. Second, the teachers were considered to occupy a strategic position in the communication processes to be explored by the studies, and could be expected to be relatively more verbal on these issues than the population average. Third, there was considerable interest among the participating teams in the conditions and attitudes of teachers *per se*. Thus, there was much motivation to use this interviewing opportunity to collect comparative data on the professional problems and opinions of this important occupational group.

Details on the samples taken in the seven countries are given elsewhere in this issue.[3] Suffice it to say that the divergencies which had to be introduced in the sample designs in the different countries as well as the marked differences in sample loss and replacement procedures tend to reduce the value of direct comparisons between national response distributions on single questions or items. The findings to be reported in this paper are all based on analyses of the directions of relationships in cross-tabulations of two or more responses.

Despite the many serious difficulties inherent in an international project of this kind (not excluding floods), it proved possible to complete the field operations in all countries by May, 1953. Each national team handled the coding of its own interview materials on the basis of common core codes set up centrally, but with the collaboration of all the teams. Each team transferred the results of coding to IBM cards, for processing by the Sociological Research Center at the University of Nijmegen, the Netherlands, and by the Institute for Social Research in Oslo. The Center at Nijmegen did yeoman work in reconciling a multitude of technical differences in coding and punching practices among the various teams. The comparative analysis has thus far been concentrated on the relationships between perceptions and opinions relating to international developments and attitudes on matters of national policy. The findings to be reported in this paper are based on some preliminary work with the data in these areas. Independent national analyses concentrating on the teacher-specific data—job satisfaction, status perception, opinions on educational problems, etc.—are in progress. A particularly detailed analysis of the English data has been undertaken by Dr. Hilde T. Himmelweit at the London School of Economics and will be published elsewhere.

Concepts

The initial research model agreed upon by the team representatives for the two parallel sets of comparative studies directly related variation in "threat" to variation in "pressures for conformity." This was a model most suited for use in the investigation of interaction processes at the face-to-face group level. The task of the attitude survey teams was to explore the possibilities of its adaptation for research at the national and international

[3] *Op. cit.*

levels. In face-to-face group situations it would plausibly be hypothesized that, within given ranges of variation, increased threats to the values of the group would lead to increased tendencies to conform to common group policies, increased pressures on others to conform, and increased rejection of deviants from group policies. The problem was: would it make sense to start out from the corresponding sets of hypotheses in accounting for relationships within the nation as a unit?

It was obvious from the outset that only a modest beginning in the direction of a clarification of this problem could be made through our pilot investigations, but the task was considered a challenge worth some serious efforts. The differences between the face-to-face group situation and the national context were only too obvious; nevertheless, evidence from historical analysis, as well as speculations related to theories of intergroup conflict, would seem to lend some plausibility to the notion that similar relationships between threats and pressures for conformity might apply in both cases. It was clear, however, that in the actual development of the research designs, the terms of these relationships—"threat" and "pressures for conformity"—would have to be given very different operational referents and the relationships themselves modified and made more complex to take intervening factors into account.

The crucial difference between the group experiments and the attitude surveys was seen to lie in the conceptualization of "threat". In the group experiments, two aspects of "threat" were combined to constitute the "independent variables": the *probability* of the occurrence of events endangering the attainments of a group goal, and the *valence* for the group of the goal thus endangered. These aspects of "threat" were manipulated to produce objectively different conditions for group functioning; the success of these manipulations was tested on attitudinal and perceptual data, and the effects of the different "threat" conditions were observed and categorized in terms of pressures for conformity and aggressive behaviors toward deviants.

In the study of corresponding relationships on the national level, *objective* differences in the threat conditions could not enter into the design as such. The study was confined to Western Europe, and it would make little sense to try to establish differences between countries or between regions in this respect. The sample survey inquiries had to be squarely focussed on differences in the ways threats on the national and international levels were *perceived* and *responded to* by the subjects: what features of the international situation were most worried about, what developments in world affairs did they consider the most likely, what forces did they see at work in these developments, which nations would they hold responsible if the situation developed into a new world war? Four major aspects of this perceptual complex were distinguished in the design for the attitude study: the *salience* of the subject's worries about international developments, the *probability* of the occurrence of different international developments, the *valence* of

alternative developments, and the *perception of the principal sources of dangerous developments.*

To avoid what could be endless complexity, only a limited number of the possible international developments were explicitly taken into account. For the measure of salience, the differentiation was simply between developments which would or would not lead to a new world war. The same was true for attribution of blame or responsibility. For the measures of probability and valence, five general lines of development were considered: continued "cold war", increasing influence of the United States in Western Europe, increasing influence of the Soviet Union in Western Europe, rapid development toward a new world war, gradual development toward really peaceful conditions. Subjects were asked to order these developments in terms of their likelihood from most probable to least probable and in terms of their valence, or desirability from "best" to least good". (Open-ended questions in other parts of the interview were added to check the reliability of these orderings). "Salience", "probability", "valence" and "blame", to give them simple designations, then, constituted the primary aspects of "threat" in the attitude studies and served as points of departure in the analysis of hypothesized relationships.

Returning to our contrasts, in the group experiments the "*dependent* variables" were defined as three kinds of manifestations of pressures toward uniformity: (a) tendencies to change one's own opinions and behaviors to conform to those of other group members, (b) tendencies to change others, (c) tendencies to reject those who persisted in maintaining deviant opinion, and thus to redefine the boundaries of the group. In the parallel interview studies of attitudes to national and international problems, what one hoped were a very roughly similar set of manifestations, were to be related to differences in threat perceptions: (a) the respondent's *opinions of his government,* its domestic policies, defense efforts and general management of foreign affairs, (b) his *perceptions of national unity* and *disunity,* (c) his *optimism-pessimism* with regard to *possibilities of influencing national policies,* and (d) his views on *dissent* and on *deviants from national policy.*

Situational Factors

In this connection, an important point to be noticed is that in the interview studies, these manifestations of "pressures toward conformity" had to be specified within the framework of the *actual* political situation in each of the seven countries: the actual policies of the governments, the amount of support for these policies in the publics, the general relationships between majorities and minorities on these issues, the current situation for different types of deviants in the countries. Indeed, the specifically different societal framework in which *teachers* as a subgroup live within each of these nations are of relevance. These circumstances increased very much the importance of cross-national comparisons of findings, for in a way the historically given differences in social organization, policies and opinion structures

among the seven countries entered as additional elements to modify some of the hypothesized relationships in important respects.

The world being what it is and not as social scientists would have it, these countries did not lend themselves to inclusion in clearly distinguishable cells of a neat design—but they certainly presented marked differences in their foreign and domestic policies. To take foreign policy first, four of the seven countries (Belgium, the Netherlands, Norway and the United Kingdom) had joined the North Atlantic Treaty Organization and found strong support for NATO policies in public opinion; a fifth country, France, had also joined NATO but its public opinion was profoundly split over the policies pursued in this body. A sixth country, Western Germany, had still no foreign policy of its own, but its government had wide popular support for its decision to join the Western alliance system. The seventh country, Sweden, had decided to stay neutral in the current power conflict despite wide-spread popular sympathies with the West, and despite active advocacy of NATO-oriented policies among liberals and conservatives. On domestic policies the countries also differed markedly: there were conservative or center party governments in Belgium, France, the United Kingdom and Western Germany, a coalition government with strong pro-labor elements in the Netherlands, a labor-agrarian coalition government in Sweden and an all-labor government in Norway. These differences in political situations could obviously be expected to be reflected in differences in the response distributions for the samples of teachers in the seven countries; in the individual country data, they could be expected to modify the interrelationships between the threat perception responses and the responses relating to political conformity.

In sum: this dependence on historically given situational factors reflects the conditions and contexts of data-gathering in sample inquiries into individual attitudes. In the group experiments, the crucial conditions for the interaction processes could be more or less controlled, and the processes themselves observed and categorized. In the interview studies, the entire design had to be fitted directly into the concrete historical situation; the data gathering had to be confined to the registration of individual differences in response patterns. These response patterns were end-products of interaction processes in themselves beyond the reach of the research methods used.

Potentialities and Limitations of the Cross-National Surveys

It would seem, then, that the attitude surveys could not be assumed to examine the *general effects* of threats to a nation on the behaviors and opinions of its members in the way the experiments would in the case of face-to-face groups. We nevertheless believe the interview studies to be useful for exploring, as cognitive patterns, the relationships between different ways of perceiving possible threats to the nation and different views of national policies, national unity and national dissent. These perceptions and

these viewpoints *may* reflect interaction processes in the primary groups in which each subject is involved, but they are also responses to the great flow of stimuli coming via the mass media of communication. Essentially, what the attitude studies could help to uncover were patterns of responsiveness to major themes, appeals and recommendations in the current flow of mass-directed communications relating to national policy, defense and foreign relations.

These patterns of necessity reflect differences among countries and political sectors in the general tendencies of such communications, but they can also help to underscore some basic similarities in the modes of argumentation on these issues. Such relationships as can be empirically established between threat orientations and views about national policies, national unity and national dissent, need not necessarily reflect any dynamic interaction processes of the kind observed on the face-to-face group level. These relationships might perhaps be more adequately explained from a consideration of the predominant patterns of appeals and recommendations related to such issues.

Theorizing and research on propaganda and persuasive argumentation seem to warrant the assumption that references to threatening developments in mass-directed communications will normally be combined with "reassuring recommendations" of policies designed to avert the dangers. The threat references, as important functional elements in continuous processes of persuasion, tend to justify or facilitate acceptance of existing or proposed policies or actions. In a sense, therefore, the comparative interview studies might be said to examine the extent to which those who respond to threat appeals in some specified way respond in a similar way to the policy recommendations regularly related to these appeals in the mass media.

In the countries of Western Europe the crucial issue in international politics has, at least since 1948, been centered about the threat of Soviet aggression. Public debate on foreign policy issues has been primarily concentrated on the assessment of this threat (including the question, does it exist at all?) and on the measures to be taken to reduce it or avert it if it exists. In six of the countries in which surveys of teachers were made, the prevalent patterns of mass-directed argumentation have combined references to strong Soviet threat with appeals for a firm and powerful alliance of the Western powers, increased defense budgets in each of the countries, strengthened national unity, and isolation of dissentients from basic national policies. We would call this the *"enemy-oriented" pattern* of persuasion. In Sweden, the governmental policy was for neutrality in the current power conflict, but otherwise the predominant patterns of argumentation in the mass media were substantially the same as those in the other six countries.

Against the "enemy-oriented" patterns considerable counter-forces had manifested themselves and had been gaining strength in several of the countries, particularly in France, during the two or three years preceding

our interviews. These counter-appeals and counter-arguments took highly divergent forms, but they all attempted in one way or another to decrease the persuasiveness of the threat appeals, the policy recommendations and the correlate opinion pressures predominant in the "enemy-oriented" patterns. The major elements in these counter-patterns would seem to be: (a) emphasis on the dangers of a world war *as such,* rather than on the threat from a potential enemy; (b) emphasis on the dangers involved in the current tendency to rely primarily on military preparedness efforts in working for peaceful international relations; (c) emphasis on the need for a general reduction of tensions through patient negotiation and through increased support for neutral forces mediating between the major powers. Correlated with this *"war-oriented"* pattern of argumentation, there seemed to be a tendency to assume critical attitudes on issues of domestic policy, and a tendency to defend the rights of opinion minorities.

This contrast, then, between "enemy-orientation" and "war-orientation" in current public discussion in Western Europe could be expected to find reflection in the relationships among interview responses. A prime objective in the analysis would be to explore the consistency of the threat perceptions and the extent of their correlation with viewpoints on national policies. A secondary aim (and a very important one, if some of these relationships should prove to be consistent in the data), would be to investigate the social and personal characteristics of the respondents found to differ in their threat perceptions.

It was of course impossible to include in the design of this set of studies many of the factors that might be relevant in explaining these differences among respondents. The factors that were included primarily related to the specific personal and occupational problems of the population category to be sampled: the primary and secondary school teachers in each country. These included: the usual demographic and social data, measures of general job satisfaction, perception of teachers' social status, atttitudes to professional problems, general views on discipline and principles of behavior. In this way the studies also provided comparative data of intrinsic value for the understanding of the positions and problems of the teachers as an occupational group in the different countries. A report on this teacher-specific phase of the study program may be prepared later.

Analysis

Let us turn now to some of the early results of the analytic work that is in progress, and to some of the problems of their interpretation. The relationships among the various responses involved in threat perception will be considered first, then the relationships of threat perception to views on government policies, national unity and dissent. What has been *sought* in the analysis is suggested by the very brief conceptual framework that already has been presented; what has been *found,* of modest substance, is reported with the hope that it be seen as part of the OCSR's methodological learning process, as exploratory behavior in the latter part of the research maze.

Threat Orientation

How are the different responses to aspects of possible threat in international affairs related? Four such aspects and responses were mentioned earlier:

(a) the *salience* of the respondents' worries about world affairs;
(b) the relative *probability,* as respondents saw it, of the occurrence of five possible developments in Western Europe—war, American dominance, Russian dominance, continued cold war, genuine peace;
(c) the *valence,* or relative acceptability and unacceptability to the respondents of these same five developments, and, in particular, war;
(d) the location or *source* of dangerous developments seen by the teachers in the samples, and consequent attribution of blame.

The principal finding is that only one of the possible pairs of these responses is consistently associated in all the seven countries. The *valence* of war, relative to other possible international outcomes, is closely related to which of the major power constellations is seen as the *source* of danger in the world.

The *valence* measures were two: (a) the ranking of the five possible developments from "would result in the *best* situation for us if it should happen" to "would result in the *least good* situation for us", and (b) responses to the question: "Under what circumstances would you look upon a world war as an alternative that would be the 'lesser of two evils?' " In all seven countries, those teachers who ranked the outbreak of war as the "least good" possibility were significantly more likely to say that under *no* circumstances can war be a "lesser of two evils".

The *"blame attribution"* or *"source of danger"* measures were based on content analyses of responses to two open-ended questions: "In your judgment, what developments in international affairs constitute a danger?", and "If a world war were to occur, do you think any one nation or group of nations would be most to blame?" Those who referred to the Soviet Union and the Communist bloc as the principal source of danger also tended to attribute the blame for a new world war exclusively to these powers; this is the case in all of the countries.

When the valence and "blame attribution" measures are examined jointly, we find consistently that those who *could* conceive of circumstances under which a world war would be the "lesser of two evils" were more likely to put the blame exclusively on the Soviet bloc, while those who *could not* conceive of any such circumstances were less likely to blame the Soviet bloc and more likely to stress the general features of the bipolar power struggle or, more rarely, the responsibility of the Western powers: these relationships will appear from inspection of Table Ia. The consistency and significance of this relationship in all the seven countries give us some reason to believe that the basic distinction in orientation to world events that we discussed earlier is actually reflected in the response patterns of the teachers—the distinction between an *enemy-oriented* emphasis,

TABLE I

VALENCE OF WAR, BLAME ATTRIBUTION, AND APPROVAL OF DISSENT

A. *If a world war were to occur, do you think any one nation or group of nations would be most to blame?*

B. *Do you think it is good or bad to have fundamental differences of opinion, within a country, on such matters as military preparedness?*

Under what circumstances would you look upon a world war as an alternative that would be the "lesser of two evils"?			A. Blames Eastern bloc	A. Doesn't blame Eastern bloc	A. Don't know or no response	B. Good	B. Both good and bad	B. Bad	B. Don't know or no response	Tot	N
	Belgium:	1	35%	51	14	34%	10	46	10	100%	166
		2	57%	33	10	32%	12	50	6	100%	206
		3	—	—	—	—	—	—	—	100%	28
	England and Wales:	1	29%	46	25	49%	13	32	6	100%	161
		2	47%	29	24	52%	10	29	9	100%	210
		3	—	—	—	—	—	—	—	100%	29
	France:	1	16%	52	32	51%	16	31	2	100%	257
		2	40%	39	21	49%	11	39	1	100%	72
		3	—	—	—	—	—	—	—	100%	21
1. Under *no circumstances*	Netherlands:	1	47%	42	11	32%	16	48	4	100%	81
		2	70%	20	10	30%	7	63	—	100%	294
		3	—	—	—	—	—	—	—	100%	25
2. Under *some* circumstances	Norway:	1	26%	42	32	38%	9	43	10	100%	113
		2	57%	16	27	30%	14	53	3	100%	250
		3	—	—	—	—	—	—	—	100%	37
3. Don't know or response not ascertained	Sweden:	1	60%	33	7	14%	7	76	3	100%	139
		2	78%	18	4	11%	6	82	1	100%	276
		3	—	—	—	—	—	—	—	100%	28
	Western Germany:	1	23%	46	31	71%	11	15	3	100%	190
		2	49%	38	13	67%	7	23	3	100%	155
		3	—	—	—	—	—	—	—	100%	20

focussed on the aggressive intentions of the potential enemy, and a *war-oriented* emphasis, focussed on the danger of war as such. This distinction has proved of particular value in the further analysis of the interview data and is of central importance in the study.

It could perhaps have been expected that the war-oriented teachers would on the whole be more consciously worried about the international situation and also would assign a higher probability to the possibility of a third world war. Neither seems to be the case. In most countries there is no significant association between measures of the *salience* of international tension on the one hand, and measures of blame and valence on the other hand. In France and the Netherlands the war-oriented are significantly more concerned with the international situation, while in Belgium there is a tendency in the opposite direction: those who blame the Soviet bloc are more likely to manifest worry about world developments in their responses to our open-ended questions.

A question might reasonably arise: are these threat orientations related to the more obvious "background" or "demographic" characteristics of the teachers? In general, this must be answered in the negative, except for one critical characteristic to be discussed later. Thus, there are no constant differences between primary and secondary school teachers in their threat orientations. In Belgium and France the primary school teachers are on the whole more likely than the secondary school teachers to attribute high negative valence to the possibility of a new world war; while in the Netherlands and in Norway the secondary school teachers tend to rank higher in war-orientation. Analyses of the distributions of "blame attribution" responses also show some differences among the countries: in Belgium, England, the Netherlands and Norway primary grade teachers are more likely than secondary grade teachers to put the blame exclusively on the Soviet bloc; in the other countries, there are no clear relationships. The differences between men and women teachers may also be of some interest. In Belgium, England and the Netherlands the male teachers tend to be more war-oriented than the females, while in Sweden the opposite appears to be the case. In the other countries these relationships are unclear and unsystematic. Similar country differences are also found when the "blame attribution" responses are considered. Among the other background factors which have been treated are the teacher's age and the location of his school; the results, as in the examples above, show no distinct patterning.

Threat Orientations and Views of Governmental Policy

The second phase of the analysis considered in this paper concerns the relationships between threat orientations and views of governmental policy. Perhaps the simplest summary of these views would lie in the respondents' reports on their political party sympathies—if one could only measure these sympathies by simple means. The proportion of teachers for whom this information was not obtained in these surveys ranged from 4% in Sweden

to 19% in Germany. There were also marked differences in the proportion of respondents reporting *no* party affiliations or sympathies: the range was from less than one per cent in Sweden to as much as 31% in Germany. We find significantly more orientation to the Soviet threat among those sympathizing with the conservative and the center parties and significantly more orientation to the dangers of war among those who report pro-labor, social-democrat or socialist leanings.[4] This was the case in all countries, whatever the politics of the government then in power.

An interesting check on this association is provided by relating threat orientations to responses to the question: "Do you feel that our government is doing all it possibly can to maintain peace in the world?" In all countries within the Western defense system the enemy-oriented teachers more frequently gave affirmative replies to this question. In Sweden the tendency is the opposite: the enemy-oriented teachers were no more likely to express confidence in their government's policies for peace than the war-oriented, presumably because so many of the enemy-oriented would have preferred to see their country abandon its neutrality and join the North Atlantic Treaty Organization.

A related finding can be reported from the analysis of the association between the respondents' threat orientations and their confidence in the domestic policies of their governments. In all the countries that were governed by conservatives or Christian center parties, the enemy-oriented teachers generally tended to express more confidence in the current domestic policies than the war-oriented teachers, while in three countries that had more labor-oriented governments the tendency was in the opposite direction, although significantly so only in Norway. It would seem, then, that high enemy-orientation need not necessarily predispose toward acceptance of total national policy, as a simplicistic application of much social-psychological theory might have it; the over-all structure of domestic politics appears to be highly important in this relationship.

Substantial support for this interpretation is found in an analysis of the over-all relationship between the party sympathies reported by the teachers and their responses to the different questions used to elicit opinions on the national government and its policies. In England, France, and Germany the Labor-supporting teachers were clearly least likely to express support for their governments' foreign policies. In Belgium and the Netherlands, the Socialists were definitely less often in favor of the current foreign policies than the Catholics, but there were no clear differences between the Socialists and the Liberals in Belgium or Socialists and Conservatives in the Netherlands. In Norway, those who reported Labor sympathies were slightly more frequently in favor of existing foreign policies than those with other party orientations, while in Sweden there were no substantial differences

[4] The number of teachers who reported Communist party affiliations was much too small in all countries to allow any statistical analysis.

between teachers with different party affiliations on this score. Relating party sympathies with responses to questions about domestic policies, we find a much more clear-cut contrast among the countries: in Belgium, England, France and Germany, the Socialist supporters tended to express least confidence in the current governmental domestic policies; in the Netherlands, Norway and Sweden, they tended to express the most confidence.

These contrasts also manifested themselves in the relationships between threat-orientation and views on defense efforts. In all countries, without exception, the enemy-oriented teachers were significantly more often in favor of increased military preparedness than the war-oriented. However, if we look at the relationship between views on domestic policies and views on defense efforts, we find a major difference between the countries. In the four countries with conservative or Christian center governments, those who expressed confidence in their government's domestic policies were on the whole more likely to favor increased defense efforts. In the three other countries the tendency is in the opposite direction, particularly so in Sweden where those who disliked their government's domestic policies were significantly more often in favor of increased military preparedness.

Threat Perceptions and Views of National Unity and National Dissent

These between-country differences clearly point to the importance of the national political structures—what they are, how they are perceived and the ways they may affect views of unity and dissent. One hypothesis we examined held that the enemy-oriented teachers would in general tend to perceive wider national consensus on central issues of foreign policy and defense. This does not appear to be the case in many of the countries.

The question used to get at this perception of national consensus followed some discussion of defense policy, and was: "In your judgment, about what proportion of your countrymen agree with the government on the defense issue?" Cross-tabulating the responses to this question with the threat orientation responses, we find that only in France did the enemy-oriented more often than the war-oriented see greater national consensus. This relationship is less apparent in Germany, the Netherlands and Norway and not clear at all in Belgium and England. In Sweden, the tendencies are on the whole in the opposite direction: the enemy-oriented teachers tend to see less agreement. A more detailed analysis of the relationships between party affiliation, opinions on governmental policy, and perception of national consensus seems to clarify this finding: there was a general tendency for those who identified with the party or parties in power to see less disagreement in the nation than those who did not so identify. The exceptions for Belgium and England *might* be understood in the light of the very narrow majorities of the governing parties in these countries.

The difficulties in interpretation of relationships are even greater when we pass on from the analysis of perceptions of national unity or disunity to the analysis of feelings about agreement and disagreement with public pol-

icy. The principal questions used in the latter area were: "Do you think it is good or bad to have fundamental differences of opinion, within a country, on such matters as military preparedness?" and, "In general, what do you think should be done with people who have different opinions from those of the majority in times of war?"[5] The use of these questions has raised some important problems of cross-national stimulus equivalence; also, the questions were probably differentially interpreted by different kinds of respondents within a single country. There are significant associations between the responses on these two questions in two of the countries, less definite relationships in three of them, and no relationship in Belgium and Sweden.

If we review the way in which responses to these two questions on unity and dissent are related to other major responses in the study, we find important differences among the countries. These are differences in the way threat perceptions relate to approval-disapproval of national disagreements as such (the first question), and to tolerance-intolerance in expressions of feelings about deviant members of the national community (the second question). For example, in France the enemy-oriented respondents tended on the whole both to disapprove of disagreements and to reject deviants, but in Germany, the Netherlands and Sweden they seemed to be as tolerant of deviants as the war-oriented but more likely to disapprove of general national disagreements. Among teachers in England and Norway, on the contrary, there was less difference between the enemy-oriented and the war-oriented on the value of national disagreement, but substantial differences on the treatment of wartime deviants. In Belgium there seems to be no clear relationship for either of the measures. Table IB will give an impression of the complexity of these relationships and the difficulties of between-country analysis.

These between-country differences are currently being analyzed in greater detail. It seems to us that if these particular findings are at all interpretable a multiplicity of factors will have to be taken into account, including such material on the personality and social status of the teachers as are available in the study. Differences among our teachers in their political party affiliations and in their general acceptance of governmental policies seem to account for some of the discrepancies, but the evidence is far from conclusive. At this stage of the analysis, we believe that it will be very difficult to include the data on views of national unity and national dissent within the simple patterns of association indicated by our analysis of the relationships between threat perception responses and policy conformity responses. It may well be that our measurements in the area of unity and dissent have been inadequate for cross-national comparisons. At all events, it will not be in the least reassuring to be able to interpret some of the between-country differences in relationships more or less plausibly *a posteriori,*

[5] The French and German versions of the second question were each markedly different in content from the common item used by the five other countries, but were also intended to provide measures of "tolerance of deviants" on military issues.

from complicated considerations of the specific situations in each country and for each major respondent category.

Some Conclusions

This brief discussion of the concepts, methods and preliminary findings of the seven-nation interview inquiry will have given some indications of the complexities faced by one group of social scientists, who embarked upon systematic comparative research on the patterning of opinions and attitudes in different political and cultural settings. International communication and opinion research has made rapid progress in the years since World War II, but social scientists have only slowly come to grips with the variety and host of methodological problems and theoretical potentialities of this kind of research. If they have served no other purpose, our comparative interview inquiries have at least provided for the participants experience that may help in the clarification of the concrete research problems. We hope to share this experience with others more fully in published form.

The early analysis findings outlined in this paper give some evidence of high consistency in the patterning of threat-orientations in the Western European area and also indicate some consistent correlates of such orientations. Yet what has been most impressive in the analysis is the way in which so many of the relationships vary in size, direction and significance in different countries, the way they seem modified by specific national and international situational factors—by the historically-given structures of political forces, by the dominant policies, by majority-minority relations, by the ongoing communication processes in the mass media and in the larger organizations. To repeat an old admonition, which is especially pertinent to *cross-national* research, the interview method of data-gathering must be supplemented by other techniques and procedures for the collection and analysis of information on political forces and structures. For instance: the different relationships considered in this paper could probably be given much more adequate interpretations if in each of the countries content analyses could have been made of major trends of argumentation in the mass media, including the relative frequencies of combinations of threat references with appeals for national unity, increased defense efforts and the isolation of deviants.

The combined operations that are implied would no doubt call for very different kinds of working arrangements and perhaps some experimenting with alternative forms of collaboration than that of the OCSR. Having in mind the OCSR's varied experiences in organizing and implementing the seven-nation studies, it does not seem too cautious to recommend a more focussed approach to comparative social research. Thus, if simultaneous inquiries were required, they might be restricted to two or three countries with cultural similarities but clearly identifiable contrasts in social structure or political arrangements. The range of material sought in the interviews might be limited, but each content area thoroughly explored. The carefully

chosen samples ought to be large enough to allow extensive analyses and reliable inferences, even though interviewer training and control should be intensive enough to reduce response loss in the sample as a whole and on individual questions. Finally, the analysis of the data and the interpretation of findings might be joint enterprises among scientists from the few countries studied, and be fitted into the broadest possible framework of existing evidence in each country. Progress toward complex research arrangements of this kind will necessarily be slow and haphazard, but there can be little doubt that the task is worth the sustained effort of social scientists who want to get beyond superficially satisfactory cross-national comparisons of opinions.

Methods Used for Producing Comparable Data in the OCSR Seven-Nation Attitude Study

Eugene Jacobson

The design of the OCSR attitude study called for a set of equivalent research operations in each of the seven nations. It was clear from the outset that it would be inappropriate to work toward an ideal of identical research in every detail, both because it seemed likely that identical operations might, in some cases, result in measurements that were not equivalent, and because the variations in research facilities available in the seven nations would not allow identical operations.

The alternative adopted was to use a set of pre-tests or model studies that would permit the national teams to test the equivalence of operations that would be appropriate for the ultimate data gathering and processing in each of the countries, and, at the same time, develop more nearly similar research facilities. It was intended that at the end of the series of pre-tests the teams would be staffed and equipped in comparable fashion and that the questionnaire and content analysis procedures, though perhaps modified to meet situational country demands, would be producing equivalent data.

The preceding article[1] affords some basis for judging the success of these efforts. To the extent that it is possible to demonstrate consistency among countries in the complex relationships within the attitude data, there is evidence that equivalent data have been produced. This point will be taken up again at the end of this report. In the paragraphs that follow we will describe the steps that were taken to work toward equivalent research operation in the seven countries.[2]

The Major Planning Conferences

An initial three day planning meeting was convened by the Institute for Social Research in Oslo in May, 1951, presided over by Dr. Herbert Hyman of Columbia University, in residence at the Oslo Institute as a Fulbright scholar. Social psychologists and sociologists from Belgium, France, Western Germany, the Netherlands, Norway, Sweden and the

[1] Vilhelm Aubert, Burton R. Fisher, and Stein Rokkan, "A Comparative Study of Teachers' Attitudes to International Problems and Policies: Preliminary Review of Relationships . . .", this issue.

[2] A more complete account of the project, up to the time that data gathering began in the winter of 1952-53, can be found in *Progress Report*, Organization for Comparative Social Research, published by the Institute for Social Research, Oslo, Norway, February 1953.

United Kingdom discussed theory and methods and proposed that an attitude study be conducted in each of their countries both as a means of collecting intrinsically useful comparative social science data and as a basis for establishing uniform training opportunities in the academic institutions that the investigators represented. They decided to work on plans for such research separately during the summer and reconvene in Oslo in the fall. The four week planning conference in September-October 1951, led by Dr. Daniel Katz, University of Michigan, in residence at the Oslo Institute as a Fulbright scholar, resulted in agreement on a basic research problem, agreement on the use of an intensive interview survey method to accumulate data, agreement on a tentative time schedule and the programming of a set of initial steps for the development of research instruments, and the establishment of a central administrative office at the Oslo Institute for coordinating the research among the countries.

The third planning conference was held in April, 1952, at the Psychological Research Laboratories of the University of Amsterdam under the direction of Dr. Katz. More detailed discussions resulted in the following proposals:

1. The research problem would be to study the effect of varying conditions of experienced threat on group members as reflected in their attitudes toward deviants within the group.
2. The content of threat would be looked for in the general area of international affairs with an emphasis on threat of war.
3. The group to be studied would be primary and secondary school teachers, both because several of the investigators had continuing interests in studies among teachers and because teachers were judged to be a relatively available and readily sampled population.
4. A random and representative sample of teachers would be selected by area probability sampling where appropriate, and with the help of teacher rosters in those countries where central records were conveniently available.
5. The data would be collected by personal interviews with the selected respondents, during which they would be asked a standard set of questions, and their answers would be recorded verbatim.
6. The recorded questionnaire protocols would be coded by standard content analysis procedures to allow mass analysis with punched card methods.
7. Sub-committees of the whole seven-nation set of investigators would prepare the study instruments with the help of a project coordinator working out of the central office.
8. The project coordinator would have the responsibility of devising and administering techniques for standardizing the research operations.
9. Data would be gathered during the winter of 1952-53.

In the following sections we will discuss the steps that were taken to work out this program through the invention of techniques for approaching equivalence in each of the several major research phases.

Sampling Teachers

What kind of a sample was wanted and what was obtained?

When teachers were suggested as the subjects for the attitude study during the initial planning conferences, the implicit intention was to provide a representative sample of all of the primary and secondary school teachers in each country. With completely comparable representative sam-

ples the reports of the research could refer to the "Belgian teacher" or the "Swedish teacher". Three hundred primary school teacher interviews and one hundred secondary school teacher interviews, selected in such a fashion that they would not only be random but also that they reflect major distribution of teacher characteristics within the country, were to allow extensive comparisons among "teachers" in the seven countries. This ideal was approximated with varying degrees of closeness in the several countries.

The complementary objective, that individual respondents be selected without sampling bias and that the sampling procedures be sufficiently explicit and denotable to permit replication, was met much more satisfactorily.

What was done to provide comparability in samples?

After the first Oslo conference, a sample of Swedish teachers, stratified by geographical area and sex, was circulated among the teams as an example for discussion. France and Germany prepared alternative sample designs to meet special field conditions in their countries.

At the Paris conference it was agreed that there would not be identical samples because costs were different among countries and because already existing field organizations that had trained interviewers were to be used in France and Germany. Ultimate selection of respondent would be identical, however. That is, random selection would be made by the team leader from rosters. And the population of teachers, in each case, was defined in the same way. Only full time teachers in regular schools in the state and religious systems were to be interviewed.

Geographical restrictions

In Belgium, the Netherlands and in England and Wales there were to be no arbitrary exclusions of geographical areas. But in Norway, Sweden and France it was decided to use a sample of quasi-urban areas, defined as major population centers and regions adjacent to them. These areas included some "rural" schools, but none that were genuinely remote from population centers. In the smaller countries it could be assumed that very few teachers included would come from places that were, in the same sense, geographically remote, and that the decision to use a quasi-urban sample in the larger countries would not so much limit the comparability among countries as it would limit the representativeness within. In Germany, sampling was done from a set of selected areas that had been determined, on the basis of previous studies, to contain a population representative of Western Germany and West Berlin.

Sample Analysis

During the second field trip of the survey coordinator, final adjustments in the sample were agreed upon. Table I presents a condensed analysis of the primary school teacher sample in each of the seven countries.

Representativeness. In discussing the OCSR study data from Belgium,

TABLE I

NATIONAL TEACHER SAMPLE: PRIMARY SCHOOL TEACHERS

Specifications	*Belgium*	*England & Wales*	*France*	*Germany*	*Netherlands*	*Norway*	*Sweden*
Total no. of teachers	35,801	155,291	194,000	126,380	35,398	11,500	36,029
Geographical restrictions of population	None	None	Areas around major cities in 8 geogr. regions of France	See Text	None	Excl. 3 North counties	Areas around 4 major cities
Pop. sampled	35,801	155,291		126,380	35,398	9,725	6,345
Sample drawn	300 + 150 reserves	300 + 300 parallel samples	338 + suppl. 98	300 + replacements	300 + replacements	300 + 300 reserves	437
No. teachers interviewed	300	300	272	265	300	300	324
Sample loss: Total %	13% of 300	14% of 300	39% of 436	45% of orig. sample	9% of 300	10% of 300	26% of 437
Refusals %	8% of 300	8% of 300	29% of 436	25% of orig. sample	9% of 300	6% of 300	8% of 437

Pop.-sample Comparison	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample	Pop.	Sample
Male	44%	44%	42%	38%	31%	42%	63%	64%	54%	54%	58%	54%	34%	29%
Female	56%	56%	58%	62%	69%	59%	37%	36%	46%	46%	42%	46%	66%	71%

England and Wales and the Netherlands, the table above suggests that it would be possible to make generalizations about *"the full-time state system primary school teacher"* in these countries. There were no arbitrary geographic sample restrictions, loss was relatively low and stratification probably increased representativeness. If sampling were the only consideration, direct comparisons of OCSR data among these countries would be justified. In Norway and Sweden refusal rate is relatively low, but part of the teacher population was not sampled. Generalizations would have to be limited to *"teachers in the areas sampled."* In France and Germany there is both arbitrary geographic restriction and relatively high sample loss. The sample can be described, strictly, as a *"partial sample of teachers in some areas."* It is a reproduceable sample, but it is not representative of the teachers in these two countries in the same sense as in the other five. These considerations suggest that findings from all seven countries, when comparative analysis is presented, cannot be generalized to "teachers in these countries", but must be qualified as proposed above.

Randomness. The ultimate selection of respondents was the task of the research director, not the interviewers, in all of the countries. Interviewers were furnished either the names of the teachers or the characteristics of classrooms designated unambiguously, and had no opportunity to influence the choice of respondent. The research directors made their selections randomly from lists of teachers or classrooms both for the original sample and for replacements of sample loss. In replacement, selection was made from strata representing the loss.

Reproduceability. Individual team records will allow other researchers to duplicate the samples used.

Gathering Interviews

What interview questions should be used?

It was primarily the problem of arriving at a set of interview questions that would be acceptable to all of the teams that led to the elaborate structure of pre-tests, conferences, and memoranda distribution that the OCSR survey used. Seven principal investigators in seven countries had to be satisfied that a very concrete and vulnerable collection of questions in fact represented appropriate means for getting evidence about a relatively ambiguous framework of constructs accepted as the survey design.

After the second Oslo planning conference in the fall of 1951, Dr. Katz and the Norwegian team circulated a first draft of the questionnaire to gather data about the three major variables: threat of international conflict, response to threat represented by attitudes toward deviants, and intervening determinants of basic personality predispositions and general value systems. Before the spring, 1952, meeting in Amsterdam, each team interviewed thirty respondents. On the basis of the discussion of this experience at Amsterdam and the agreement to use teachers as subjects, Prof. Oldendorff,

Dr. Jetter, Dr. Katz and the Norwegian team prepared a second pre-test, including in addition to revisions of the original material a group of questions designed specifically for teachers. There were about two hundred questions in this form, intended to produce data related to forty variables.

In May and June the teams used this version as an intensive qualitative pre-test with ten teacher respondents. Verbatim transcripts of respondent's answers and extensive comments on the field experience were sent to the Oslo office as a basis for selecting the best items for the next draft.

The final pre-test form of approximately the desired length and sequence—about forty substantive questions in addition to background material on age, education, etc.—was used in a full-scale dress-rehearsal with 50 teachers in each country in the later summer and early fall of 1952. The survey coordinator visited all of the teams during this pre-test and, after discussions of the field experience with team leaders, issued another revision of the form to be used with ten respondents.

From the 50 interviews, team leaders sent a random sample of ten to the central office along with a complete tabulation of all 50 on a categorizing scheme drafted by the central office after review of the previous pre-test material.

All of the principal investigators met in Paris at the Centre des Etudes Sociologiques for two weeks at the end of the pre-test period in October, 1952. In the course of the conference, five additional revisions, developed after discussions of individual team experience and analysis of the data from the pre-test of 50 tabulations, culminated in a questionnaire form that all team leaders accepted as being appropriate in wording and sequence.

How should the questions be worded in the different languages?

When the Paris meeting agreed on the exact wording of the questionnaire, the consensus was about intent and meaning in the group's common language, English. But the problem of producing functionally equivalent translations in the languages of the seven countries remained.

There were two aspects to the multiple language problem: the translation had to reproduce the English meaning adequately, and the adequate translation had to be judged in terms of its psychological equivalence to the forms being used in the other countries. The operations that were intended to meet the translation problem are described in the following paragraphs. At the end of this section we will speculate about some *ex post facto* checks on functional psychological equivalence.

The languages of the study included Norwegian, Swedish, German, French, Dutch and English. Regional language differences within countries were not formally considered with the exception of the use of variations of French and Dutch as the two languages in Belgium.

Translation. In the analysis by the central office of the translations of

the initial pre-test of ten and the following pre-test of 50 questionnaires, and during the conversations of the survey coordinator with team members, a number of gross translation difficulties were resolved. The three general classes of problems and their solutions were: a.) When the team leaders and central staff agreed that there were several equally appropriate translations into the language of use, the team leader chose the form that he judged best in establishing rapport or in terms of simplicity or style. b.) If the original question was intentionally worded ambiguously, the central office reviewed the translations among the countries and recommended a wording for each that tended to establish the ambiguity in equivalent fashion. c.) In those instances where the translation difficulty occurred because the concept was not commonly found in the culture or the language was not equipped to handle the concept unambiguously, the question was eliminated in all countries. Then an effort was made to find an alternative useable concept and translation that met the study objectives.

These language adjustments were made in all of the pre-test stages. The final check was made after the Paris meeting when each team sent five independent translations into the central office, preferred forms in each language were selected and discussed with team leaders, and a compilation of official versions in all languages was issued.

Functional Equivalence. The discussion above has been about the techniques developed for arriving at a translation that seemed to preserve the intent and meaning of the original English form. The complementary question about whether these good translations would produce equivalent data among the countries has to be approached in another way. In different cultural contexts formally equivalent translations might result in questions that would produce different classes of data. Or, the set of adequate translations of what was a good sequence of questions in English might produce a badly ordered set of stimuli in the language of use.

The extensive pre-testing was intended to uncover problems of this sort, and it is the judgment of the team members and the central office staff that many equivalence difficulties have been avoided. But the evaluation of the success of this aspect of standardization has to be done during the analysis of data. In general, the evaluation will rest on the extent to which sets of questions among the countries are capable of providing data that allow the examination of parallel relationships. For instance, if in one country attitude "x" is found to be related to attitude "y" and attitude "z" and background data "d", the same set of relationships should be testable in the other countries. If the data are available and distributed in such a fashion that these relationships can be tested in all countries, one item of necessary, but not sufficient, evidence about equivalence has been found. If the same relationship is found among countries to a degree not attributable to chance, a very convincing demonstration of the equivalence of questions has been established. As a greater number of interdependent systematic relationships within sets of data in one country are replicated in

other countries, the evidence for equivalence becomes stronger. Examples of this are found in the preceding article.[3]

Similarly, systematically different sets of relationships within data among countries, when these systematic differences have been predicted on an *a priori* basis, constitute good evidence of equivalence. The unresolved problem is the interpretation of the absence of systematic similarities or variations within data among countries. This could be the result of unanticipated cross-cultural variation or of simple methodological failure of a number of kinds as well as failure to maintain functional equivalence in translation. In general, negative evidence in this kind of investigation is very difficult to interpret. A complete evaluation of the functional equivalence of questions that will include negative evidence will be possible only when the properties of the variables being investigated are more fully known than those that are used in this study.

The methodological problems of language equivalence in cross-cultural investigations are being studied by Mr. Stein Rokkan and will be discussed at length in reports from the Organization for Comparative Social Research.

How were the interviewers selected and trained?

With a standard interview and controlled translation, there still remained the problem of providing an interviewing situation in each of the countries that would be equivalent. The final draft of the questionnaire included a sequence of 39 questions designed to produce extensive answers and 18 additional items about the respondent's teaching experience and personal background. In this type of almost clinical interviewing—each interview took from one to two and a half hours or longer on the average—under a variety of field conditions, the standardizing of interviewer behavior becomes very important.

Selection. Interviewers were either graduate students in the social sciences or trained professional interviewers on the staff of continuing field organizations. In Belgium, the Netherlands, Norway and Sweden the team leaders recruited graduate students in sociology, psychology, economics, law, anthropology and related disciplines. In France, interviewing in the Paris area was done by graduate students, and in the other areas the professional field staff of the national government special surveys unit did the work. In England both graduate students and employees of the national social surveys group interviewed, under the supervision of the team leader. In Germany a trained group of interviewers who had been doing studies for the survey organization in which the team leader was a research director were available. A few of these persons were graduate students. To complete 300 interviews in the ten week field period, centered around January-February, 1953, the teams used these field staffs: Belgium—37; France—21;

[3] Aubert, Fisher, and Rokkan, *op. cit.*

Germany—46; Netherlands—51; Norway—30; Sweden—38; and United Kingdom—32.

Training. Interviewer training was continuous during the three pre-tests and depended heavily on reports of field experiences. To begin the process, some of the teams that had done intensive, quasi-clinical interviewing in field situations, distributed some training materials to the other teams. When team comments on the first pre-tests were available, this material was revised and incorporated into a version of University of Michigan Survey Research Center interviewer training manual adapted for the OCSR.

The pre-test of 50 questionnaire was accompanied by extensive item-by-item analyses of question objectives as training aids. This material was revised after the pre-test of 50 had been discussed at Paris and was used in a final summary of detailed instructions.

Content Analysis

Content analysis is still more of an art than a science when it is used within nations. The seven-country uniform coding requirement, with the additional burden of among-nation variations in language and culture, could be met only by attempting to arrive at common practices through a series of approximating operations.

Code Preparation

Most code categories for questions aimed at major study objectives were implicit in the way in which questions were worded. For instance, the question "Are developments making another world war more or less likely during the next two or three years?" was intended to produce answers that could be coded on a five point scale from "much more likely" to "much less likely". Analysis of pre-test interviews suggested additional categories. In the course of the pre-tests, conferences and team visits a draft of a code was developed for use by the time that the dress-rehearsal pre-test of 50 was in the field. Interviews from this final pre-test were coded, the data and the categories discussed at the Paris conference, and a revised working code was ready for use during the final field interviewing. With some supplementary categories demanded by the full-scale interviewing, this fourth draft was issued as a definitive code.

Coder Selection and Training

In most of the countries the coders were graduate students in the social sciences, many of whom had worked as interviewers in collecting the data. In all countries they were selected and trained by the team leader and the principal staff assistants. The training in most countries included a discussion of the study, instruction in the use of the code, and a set of practice sessions in which all of the coders worked with a common set of interviews, comparing coding assignments until there was agreement on the use of the code. Standard practices were adapted from the coding guide used by

the Michigan Survey Research Center. The number of coders in each of the countries was: Belgium—4, France—10, Germany—5, The Netherlands—25, Norway—15, Sweden—10, United Kingdom—10.

Coding Reliability

There were two approaches to the evaluation of coding reliability. In each of the countries, during the coding process, the team leader and principal assistants set up check coding procedures to control and correct the error during the actual coding. This was done, in most cases, by having a sample of the interviews coded independently by more than one person and resolving differences in discussions with the coders. In some of the countries records of discrepancies were kept from which estimates of coding reliability can be made. Such data were not accumulated by the central office.

The second effort to check on the comparabilitv of the coding operation was made when each team was asked to code five Norwegian interviews. All five of the interviews were coded by each of four or five regular coders, and the data were sent to the central office for analysis. From these data it was possible to make some judgments about the relative coding difficulty of elements in the study and to determine whether there were gross problems of code interpretation among countries. A preliminary examination of this material did not reveal gross discrepancies, although it did disclose varying degrees of agreement among countries and among items.

Preparing Data for Comparative Analysis

When the coding of the interviews had been completed in all countries, each team leader had the data transferred to punched cards. The card column numbering and assignment of codes to column positions was specified in complete detail in the code. In each country there were some idiosyncratic practices of data preparation because data were not identical and some discrepancies in processing were inevitable. But for seven-country analysis it was necessary that a common core of formally comparable data be assembled.

At the Paris conference this task was assigned to the Netherlands team. Professor Oldendorff set up a data reconciling unit at Nijmegen under Drs. Brackel. Each team sent its punched cards and its final code to Nijmegen, where a master code and a master set of cards were prepared. In the course of this reconciling operation a systematic error check was introduced and the inevitable coding and punching errors were referred back to the countries for correction.

All of the teams have access to all of the cards and the master code, although complete sets of cards are in existence only in the central office in Oslo and at Nijmegen.

Coordinating Analysis

Two primary kinds of agreement had to be reached to allow effective

exploitation of the data on a cross-national basis. First, there had to be consensus about the assignment of given items of data as appropriate evidence about the principal theoretical constructs. And second, there had to be a division of labor on the use of data in working through the elements of the analysis.

Usefulness of Coded Data

Consensus about the usefulness of given items of data rested on these agreements: 1.) In the planning stages, questions had to be perceived as being capable of producing appropriate data. 2.) When the data were accumulated and reduced to code categories, the data had to be perceived as being useful in the analysis. At every point in the development of the questionnaire and the code, an attempt was made to key the questions back into the original theoretical scheme, and there was exchange among the teams about the appropriateness of using a given item to get data about a particular theoretical construct.

By the time of the Paris conference, when data from the pre-test of 50 were available, it was possible to make some judgments about the expected usefulness of various coded items. They were examined for relative proportion of "not ascertained" categories where the question had failed to produce codable evidence. They were examined for relative frequency of "don't know" responses to determine whether this was an interpretable datum or evidence of failure of the instrument. Frequency distributions were examined to determine whether there was enough variability to allow analysis.

All of these considerations again were used at the London conference when the data from the final field operation were available. Tabulations of distributions on all items for some of the countries were discussed in detail by all of the team leaders and a tentative elimination of items was made.

When the master code had been prepared, a complete tabulation of data for the seven countries was issued. Review of this material among the teams resulted in the selection of fifty codes for the analysis of twelve major attitudinal variables. These included perception, importance, extent and source of threat of war; attitudes toward deviants and deviance; evaluation of respondents' effectiveness in influencing international affairs; basic philosophy of relationships among teachers and pupils, attitudes toward nation and government, and attitudes toward teaching and the teaching situation. Another thirty items of background material, including occupational history, age, sex, etc., and some material designed for specific nation interests brought the total of recommended code items to about one hundred to meet the study design requirements set up in the initial Oslo planning sessions.

Dr. Burton Fisher, of the University of Wisconsin, was in residence at the Oslo Institute as a Fulbright scholar during the 1953-54 period when these data were analyzed and prepared for publication.

Division of Labor

There are three kinds of analysis to be done with this material: 1.) The basic hypotheses of the study are being worked through on a seven-country basis. 2.) Within countries, nation-specific analyses are being done. 3.) When specific country interests develop analyses that seem appropriate in other countries, the multi-nation comparison will be undertaken.

Interpretative and Methodological Problems of Replicated Research

Stanley Schachter

The present paper discusses and illustrates some of the interpretative and methodological problems involved in replicated cross-national research by describing limited aspects of the set of replicated experiments conducted by the Organization for Comparative Social Research in Belgium, England, France, Germany, Holland, Norway and Sweden. The results of these studies have been presented in detail elsewhere.[1] Rather than repeating this material, we shall attempt to consider a single, key question of replicated research, "If there are differences in the results of replicated experiments, how are such differences to be interpreted?"

Though systematic replications have been unfortunately rare in the social sciences, a sufficient number of such studies have accumulated to give us some experience with the interpretative problems involved in replication. When the results of replicated researches have been similar, the interpretation is usually clear and straightforward. Reproducibility has been demonstrated and, if the subject populations or the design of the experiment has been allowed to vary, there is evidence for the generalizability of the hypotheses involved. When the results of replicated researches have differed, however, interpretation has usually been exceedingly ambiguous. Occasionally, the researchers involved have been in a position to evaluate and understand precisely the cause of differences; more frequently, a difference in results has precipitated a major interpretative dilemma (e.g. recent controversies in the areas of social perception and latent learning).

For any specific difference between experiments any of a number of reasons may serve as a plausible explanation. The differences may simply be attributable to chance. Or, they may be due to experimental artifacts, such as failure to properly manipulate the independent variables. Or, such differences may genuinely reflect cultural differences among the subject populations. Or, differences may indicate that the original hypotheses underlying the experiments may have been incorrect or only partially correct, and an alternative formulation may be able to reconcile apparent divergent results. For many differences between experiments, any of these may be a possible explanation. It is clear that if a particular set of replications is to be more than a vehicle for potential bickering, it must be possible, in some degree,

[1] Schachter, S., Nuttin, J., de Monchaux, C., Maucorps, P., Osmer, D., Duijker, H., Rommetveit, R. and Israel J., "Cross-Cultural Experiments on Threat and Rejection," *Human Relations,* 1954, 7 (No. 4).

to determine which of these factors is responsible for a difference, should it occur.

We shall attempt in this paper to document the above generalities and to describe the kinds of precautionary procedures which can, in part, reduce the ambiguity of interpretation of experimental differences should they occur. To provide a concrete reference for consideration, the experiments conducted by the Organization for Comparative Social Research will be described briefly.

The OCSR Group Experiment

In each country an identical experiment was conducted. Reactions to a deviate who blocked progress to a group goal were studied in the experimentally produced setting of a threatened group goal. The desirability, or valence, of the goal, and the probability of reaching the goal were chosen as independent variables and so manipulated as to produce four experimental conditions:

high valence; low probability of reaching goal
high valence; high probability of reaching goal
low valence; low probability of reaching goal
low valence; high probability of reaching goal

In the various countries there was an average of between 7 and 8 groups in each condition.

The experiment proper was conducted in the guise of the first meeting of a boys' aviation club. In the several conditions, valence was manipulated by establishing goals of varying desirability. Probability was manipulated by instructing some groups that they had a very good chance of reaching their goal and other groups that they had a very poor chance. In order to decide which groups would reach their goal, a competition was announced in which the groups building the best model airplanes would achieve their goals. The boys were presented with five possible models to build and were asked to decide which of these they would like to build. Four of the models were extremely attractive, motor-driven planes; a fifth was an uninteresting glider model. One member of each group was a trained confederate who always chose the glider. The remaining members almost always chose one of the other four planes. In each group, the subjects had a twenty-minute discussion of which plane to choose. During discussion the subjects usually reached agreement among themselves as to which of the four attractive planes to build. The confederate maintained his deviate choice of the glider all through the discussion. The discussion was systematically observed by two observers in order to provide a record of the influence process.

Following discussion two sociometric measures were administered: 1.) a social preference scale designed to measure how much the subjects wanted to have the confederate deviate in their own working sub-group;

2.) a role preference scale designed to measure how much the subjects wanted the deviate as club president.

Experiments of this sort tend to be somewhat complicated. They sometimes require several people to play exacting roles. They require training observers to acceptable levels of reliability. They demand the invention of measuring instruments which are, at least, roughly comparable in their several translations, and so on. With so many potentially complicating factors, it is more than reasonable to suspect that, despite intensive standardization procedures, should differences in the results of experiments occur, they may be attributable to differences between experimenters or observers rather than to differences between cultures or to theoretical inadequacy.

Such considerations made it clearly desirable to devise measures to evaluate the comparability and relative effectiveness of the several experiments. Specifically, the following steps were taken:

1. The success of the manipulation of the independent variables, probability and goal valence, was determined by a series of questionnaires and rating scales administered to the subjects.

2. Systematic observation of relevant remarks during the discussion was used to evaluate, in each country, the success of the attempt to create the experimental situation necessary to provide a proper test of the hypotheses; that is, a situation in which the confederate is clearly seen by the subjects as a deviate who blocks locomotion to the group goal.

3. Recordings of two complete experimental sessions were made in each country in order to permit comparison of the verbal behavior of the several experimenters and confederates.

4. To evaluate the extent to which observers in each country had common understanding of the categories in the observation schedules, a written protocol of two complete discussions was distributed to all observers in all countries. The observers coded each statement in the protocol in the categories of the observation schedules. The protocol was in English, a language which all of the observers could read with facility.

The Discovery of Artifacts

A few examples will illustrate the necessity for measures such as these. On the left side of Table I, the experiments of Holland and England are compared on the variable, rejection of deviate. Notice that in Holland

TABLE I

AN EXAMPLE OF AN EXPERIMENTAL DIFFERENCE ATTRIBUTABLE TO DIFFERENTIAL SUCCESS IN MANIPULATING AN INDEPENDENT VARIABLE

Country	*Rejection of Deviate on Social Preference Scale*			*Effectiveness of Probability Manipulation*		
	High Val. Low Prob.	*High Val. High Prob.*	*Sign. of Diff.*	*High Val. Low Prob.*	*High Val. High Prob.*	*Sign. of Diff.*
Holland	1.17	.65	.03	2.90	3.81	.001
England	.71	.67	n.s.	3.29	3.59	n.s.

the deviate is rejected far more strongly in the high goal valence, low probability condition than in the high goal valence, high probability condition. The difference is large and statistically significant. In the English experiment there is virtually no difference between conditions. One might speculate on the differences between Dutch and English culture; or, one might ask whether the manipulation of the probability variable was equally effective in both countries. The data on the right side of Table I are derived from the subjects' own ratings of the probability that they will reach their goal. The lower the score, the lower the estimated probability. The difference between conditions in Holland is large and statistically significant. In the English experiment, the difference between conditions is far smaller and not significant. The measure of probability, then, has allowed identification of the most plausible explanation for the difference between these experiments.

Another example of the kind of difference most reasonably attributable to experimental artifacts is presented in Table II. In the first column is

TABLE II

AN EXAMPLE OF AN EXPERIMENTAL DIFFERENCE ATTRIBUTABLE TO OBSERVER ARTIFACTS

Country	*% hostile remarks in discussion*	*% hostile remarks in observer comparability check*
Holland	21	10
Norway	10	2

reported the percentage of all comments made by the subjects which were coded by the observers in observation categories indicating "expressions of hostility". In Holland, 21% of all comments are coded as hostile. In Norway, only 10% of all comments are so coded. It could, then, be hypothesized that the Dutch are somewhat freer in their expression of hostility than are the Norwegians; or, one might suspect that observers in the two countries have somewhat different interpretations of the observation categories involved. In the second column of Table II, results of the observer comparability check are presented. On printed transcript of experimental discussions, the Dutch observers categorize 10% of all remarks as hostile; Norwegian observers categorize only 2% of remarks on precisely the same transcript in this category. Though the interpretation of these data is not unequivocal, there is, at least, reasonable ground for suggesting that the differences in the results of the two experiments are more an expression of differences between observers than of differences between cultures.

The Establishment of Real Differences

These two examples have illustrated the usefulness of these evaluative measures in identifying differences between experiments which are due largely to artifacts. Such measures are equally necessary, of course, to rule

out the always possible explanation of artifactually produced differences. An example is presented in the next table. On the left of Table III a comparison of rejection data in Norway and Sweden is presented. In the Nor-

TABLE III

AN EXAMPLE OF AN EXPERIMENTAL DIFFERENCE NOT ATTRIBUTABLE TO EXPERIMENTAL ARTIFACTS

Country	*Rejection of Deviate on Role Preference Measure*			*Effectiveness of Probability Manipulation*		
	Low Val. Low Prob.	*Low Val. High Prob.*	*Sign. of Diff.*	*Low Val. Low Prob.*	*Low Val. High Prob.*	*Sign. of Diff.*
Norway	4.21	5.11	.05	2.59	4.11	.001
Sweden	5.21	4.88	n.s.	2.83	3.78	.001

wegian experiment, rejection of the deviate is considerably higher in the low valence, high probability condition than in the low valence, low probability condition. In the Swedish experiment, there is a distinct though non-significant tendency in the opposite direction. Searching first for an explanation of this difference in terms of experimental artifacts, we note, on the right side of the table, that the manipulation of probability has been extremely successful in both experiments. In both countries, differences between low and high probability conditions, on the probability measure, are significant at the .001 level of confidence. Clearly, then, an explanation in terms of differential success at manipulating the probability variable is not satisfactory. On the other manipulated variables for which measures are available, there is again no difference between the two experiments. Comparison of the recorded transcript of the Norwegian and Swedish experiments reveals nothing but the most trivial of verbal differences between the experimenters and confederates in the two countries. In short, from the data and records available, it is impossible to identify any procedural artifacts which could account for this difference in results and one may, with reasonable confidence, proceed to other speculation.

Two types of explanation are worth further exploration: 1.) The differences between the two experiments may be a result of cultural differences between the two countries; or, 2.) The relationship among the variables may be more complex than was originally thought and it may be possible to construct a theory which will reconcile these apparent differences. For example, a series of figures are presented, each of which represents an alternative formulation of the relationship of the variables probability and rejection. Figure A represents the original hypothesis of the study. Rejection increases with decreasing probability. The less the probability of reaching a group goal, the more violently will the group reject a deviate who blocks locomotion to this goal. Though the Swedish data support this formulation, it is directly contradicted by the Norwegian results.

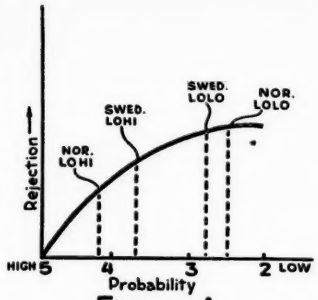

Figure A

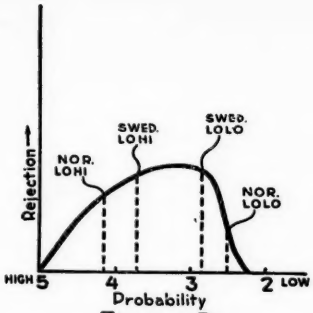

Figure B

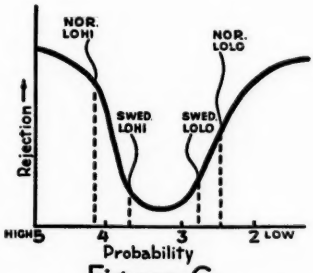

Figure C

ALTERNATIVE FORMULATIONS OF THE RELATIONSHIP OF PROBABILITY OF REACHING GOAL AND REJECTION OF DEVIATE.

Figures B and C represent alternative hypotheses as to the relationship of probability to rejection. Implicit in Figure B is the hypothesis that as the probability of reaching the goal decreases, a tendency to "give up" develops. As it becomes more and more improbable that the group can reach its goal, they lose interest in the goal and are less likely to reject the deviate.

A U-shaped relationship between rejection and probability is assumed in Figure C. Such a relationship has been derived in another paper con-

cerned with this study[2] from the interaction of "cohesiveness," "the tendency to redefine the group" and a newly postulated notion "tendency to integrate." The derivation is rather lengthy, however, and will not be repeated here.

Either of the relationships postulated in figures B and C can incorporate the apparently divergent results of the Norwegian and Swedish experiments into a coherent scheme. Along the abscissa of these figures are noted the values of the probability measure in the relevant conditions of these two experiments. If a line is projected from these points onto the curves, it will be noted that either of these formulations can incorporate these results into a single system.[3]

This example has been spelled out in order to illustrate the interpretive alternatives that remain when the explanation of differences resulting from experimental artifacts can be safely ruled out. The existing data are, of course, inadequate to provide satisfactory grounds for choosing among these alternatives. Further experiments, however, immediately suggest themselves. For example, a Norwegian medium probability condition and/or additional Swedish conditions at the low and high extremes of the probability scale could provide data indicating which of the two alternative theories was the correct one; or, might indicate that it would, at present, be more fruitful to search for explanation in the less precise realm of cultural differences.

Techniques of Attaining Equivalence

Though considerable attention has been paid to the necessity for devising measures which will allow better deciphering of the meaning of possible differences in the results of replicated experiments, it should be made completely clear that such measures are, at best, protective devices. They reduce the ambiguity of interpretation by permitting the identification of differences which are attributable to experimental artifacts. However, differences of this sort are, at best, of trivial significance. In effect such differences may mean that flatly nothing has been learned from a replication.

It is clear that such differences are to be avoided if at all possible. The execution of the experiment must be so standardized as to minimize the possibility that artifacts of the sort described will influence the results. In the study being described, efforts for standardization were particularly

---

[2] Cf. Schachter, S., Nuttin, J., *et al., op. cit.*

[3] It is possible legitimately to compare countries only on the direction of between-condition differences. The absolute magnitude of rejection scores is determined not only by probability but by, at least, two other variables as well—goal valence and group cohesiveness. The strength of cohesiveness was measured by questions such as "How frequently would you like this club to meet?" Answers to these questions were, in good part, determined by local conditions such as the number of hours the boys attended school, local labor laws, etc. Cohesiveness data, then, are not comparable from country to country and the absolute level of rejection may not be compared between countries.

strenuous. The several experimental teams represented a variety of academic backgrounds and differed widely in experience and training and without considerable effort for standardization a rash of artifactually produced differences in the results of the several experiments were feared. In order to minimize such factors, the experimenters of each nation met together several times in order to establish complete accord on experimental procedure. Manuals were prepared which contained highly detailed, step-by-step descriptions of the procedure to be followed in each stage of the experiment. A research coordinator was appointed to assume the role of traveling parameter. He visited experimental sessions in each country and attempted to eliminate variation from the agreed-upon design.

To insure some degree of language equivalence in the several measuring instruments, the following procedure was adopted. The originals of each instrument had been formulated in English, the working language of the group, then translated by each group member into his native tongue. In each country, these translated instruments were then retranslated into English by two bilingual persons, who had no knowledge of the experiment. The retranslations and the originals were compared and any differences eliminated. By this procedure it was possible to correct obvious misunderstandings and we can be fairly confident that the several versions of each measure are, at least, roughly equivalent.[4] We cannot, of course, be sure that the distance between the points of the rating scales or the location of each point on the scale are identical in the several versions of each instrument.

Nonequivalence of Situations

It is noteworthy, however, that hard work and good intentions may not be sufficient to insure proper standardization. Though it may be possible to create experiments which in execution are virtual images of one another, this alone is no guarantee that the several experiments produce conceptually equivalent situations. An effective manipulation or experimental setting in one country may be completely unsuccessful in another.

For example, in Table IV the results of the valence manipulation in several of the countries are reported. The figures are derived from a scale designed to measure the effectiveness of the valence manipulation. The lower the score the higher the valence. It is clear that the high valence manipulation was equally successful in the three countries. On low valence, however, the scores in Sweden and France are considerably higher than those in Belgium. What was relatively undesirable in Sweden and France is considered quite attractive in Belgium. The low valence goal in all three countries

[4] In order to construct instruments which would be more nearly equivalent, a preferable procedure would probably be to employ bilingual respondents to answer the original and translated form of each instrument. Comparison of responses to the two forms of the instrument would allow point for point comparison of the two instruments. Where answers to the original and translated forms of an instrument differed, the translation might be adjusted, until responses to the two forms of the measuring instrument were identical.

TABLE IV

EFFECTIVENESS OF VALENCE MANIPULATION

Country	*High Valence Conditions*	*Low Valence Conditions*
Sweden	2.04	3.50
France	2.22	3.44
Belgium	2.02	2.81

was free tickets to a rather dull documentary film. In Belgium, the subjects were parochial school boys from the Flemish part of the country and subsequent inquiry revealed that for these boys, movie going was a relatively rare privilege. A movie, then, any movie at all, would be highly valent for these boys. Evidence supporting this interpretation is found in internal analysis of the Belgian data. Some of the schools from which subjects were recruited had a more liberal "movie going" policy than did others. Boys from these schools rated the movie as considerably less valent than did boys from more conservative schools.

Another example of this sort of thing was the relative difficulty with the experimental context—an aviation club—in Germany. At the time the experiment was conducted, Germany was still under control of the occupying forces. Until a few months before the experiments all activities connected with aviation had been prohibited. The prohibition was strictly enforced and extended to such minutiae as a ban on the sale of model airplanes. The initial impact of recruiting for members for boys' aviation clubs was traumatic. Teachers were violently opposed, parents were alarmed and the Frankfurt police department launched an investigation to determine if this was an underground movement to revive militarism in Germany. With explanations the official furore died off but parental anxiety remained high throughout the time the experiment was conducted. It need hardly be added that this atmosphere was not particularly propitious for running an experiment. The subjects were ill at ease and the experimenter tense. For Germany it would have been difficult to deliberately choose a more inappropriate experimental setting.

The problems involved in producing conceptually equivalent experimental manipulations and measures are treated in detail in the following paper.[5]

[5] Rommetveit, R., and Israel, J., "Notes on the Standardization of Experimental Manipulations and Measurements in Cross-National Research," this issue.

Notes on the Standardization of Experimental Manipulations and Measurements in Cross-National Research

Ragnar Rommetveit and Joachim Israel

The present paper represents only a few general considerations based upon the set of replicated group-experiments conducted by OCSR,[1] considerations, however, that may be of some value for future undertakings of the same kind as our OCSR studies. We want to emphasize that we do not consider the difficulties to be discussed unique for cross-national studies. Rather, replication of an experiment in different cultural settings tends to accentuate some problems pertaining to empirical testing of social psychological theory in general.

"The Same Experiment" in Different Cultures

Let us first briefly consider the crucial process of translating a single hypothesis into experimental terms. Obviously the difficulties involved depend upon the way we phrase our hypothesis; the more operational its language, the easier will be its translation into experimental manipulations and measurements. In extreme cases, when concepts are defined in terms of detailed operations such translation becomes superfluous.

If our ultimate aim is to establish a body of general social psychological theory, however, strictly operational anchoring may be a very dubious procedure. (By "strictly operational anchoring" we refer to a tendency to avoid any use of a theoretical framework by reliance on operational definitions that presuppose implicit but not conceptualized assumptions. On the other hand we obviously agree that any concept should ultimately be translated into research operations.)[2] A survey of current contributions toward empirically established social psychological theory therefore by and large seems to support a more daring and flexible approach. There is an increasing emphasis on concepts (like "cohesiveness", "pressures towards uniformity" etc.) that represent abstract characteristics of social processes. The essential feature—as promising from the point of view of parsimonious explanation

[1] A full report of this research is contained in the paper: Schachter, S., Nuttin, J., Maucorps, P., Osmer, D., Duijker, H., Rommetveit, R., Israel, J. and de Monchaux, C., "Cross-Cultural Experiments on Threat and Rejection," *Human Relations*, 1954, VII, No. 4.

[2] For a more thorough discussion of flexible vs. strictly operational empirical anchoring of concepts see Rommetveit, R., "Model Construction in Psychology. A Defense of 'Surplus Meanings' of Psychological Concepts," *Acta Psychologica* (In press).

as it is troublesome from the point of view of unequivocal empirical anchoring—is the one-to-many relation between concept and observations. Thus for one single concept there may be listed a set of rules for empirical anchoring, each suggesting a research instrument by which the concept is partially determined empirically.[3] Designing *cross-national* research in terms of operational concepts may not, as we will see later on, be as good a means toward unequivocal empirical anchoring as it appears to be in other contexts.

The point of departure for our studies was a set of general hypotheses on group locomotion and rejection. They can be roughly condensed as follows: Given a group with a minimum degree of cohesiveness, locomoting toward a goal, the rejection of deviants blocking locomotion will increase in such a way that: 1.) the more valent the group goal, the greater will be the rejection of deviants; and 2.) the less the probability of achieving the goal, the greater will be the rejection of deviants. Let us select from the above condensed statement, for more careful examination, the concepts "cohesiveness", "valence of goal" and "rejection". Obviously each of the terms represents a high level conceptualization of psychological phenomena, and part of the ideal requirement for testing the hypotheses cross-nationally must read: for any given experimental condition the same (or at least, above a certain minimum) degree of cohesiveness and the same valence of goal should be established, in the different cultural settings. Furthermore, measures of the dependent variable of rejection should be comparable from country to country in the sense that the research instruments applied should assess the same dimension of rejection in different countries.

In short our task was to establish conceptually identical experimental conditions in the different countries. The range of possibilities available for us can be roughly visualized as in Figure 1.

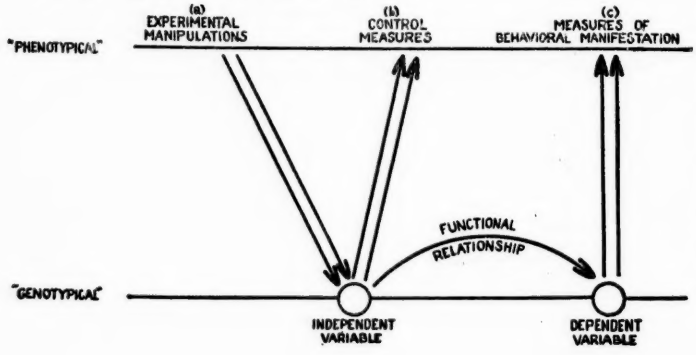

FIGURE 1.

[3] As an example, the alternative empirical anchorings of 'cohesiveness' can be mentioned. See comments on Gross and Martin's critique of the concept by Stanley Schachter in *Am. Journ. Sociol.*, 1952, pp. 555-562.

At our disposal were: a.) an almost unlimited number of possible alternative experimental manipulations; b.) a set of possible control measures for determining the degree of success of the experimental manipulations; and c.) a set of possible measures designed to assess behavioral manifestations of the dependent variable.

However, specifically cross-cultural problems of standardization enter the scheme at all points. First, divergent situational factors may affect the experimental situation so that *identical stimulus situations* fail to establish *conceptually identical situations*. (E.g. the valence of a specific goal may vary from setting to setting because of culturally determined differences of interest). Secondly, the verbal responses to control measures (e.g. to valence of goal scales and other scales used as control measures) can be affected by similar situational factors and by different verbal habits in the different countries. Thirdly, the behavioral manifestations selected as indices of the dependent variable may be significantly modified by divergent—culturally determined—past experiences of the subjects who are recruited from different settings, so that the measures obtained may pertain to somewhat different dimensions. In short, we conceive of the experiment as involving at each stage a channeling back and forth between the "genotypical" and the "phenotypical" level. The more this channeling is affected by cultural factors the more difficult will it be to establish conceptually identical experiments. These points will be elaborated in our subsequent review of a few concrete research experiences.

Experimental Manipulations and Control Measures

Our experimental studies were of the "joint development-concurrent" type.[4] As no thorough explorations of possible divergent situational factors (like systematic surveys of interests, etc.) could be made, the safest procedure seemed to be standardization at what might be called, a "verbal level". Identical goals were set up for the groups recruited from different cultural settings, experimenters were trained to behave in exactly the same way and to give identical instructions and the scales applied were translated verbatim from English to the other languages spoken in the different countries. By doing so, we tried to approach "objective" or "verbal" uniformity of experimental procedures. In short we proceeded to select one set of experimental manipulations assumed to have approximately identical effect in all seven countries.

Instead of attempting a summary evaluation of the success of our standardization procedures we shall now focus upon a few specific problems illustrating difficulties encountered at steps (a) and (b) of Figure 1, respectively.

---

[4] See Duijker, H. and Rokkan, S., "Organizational Aspects of Cross-National Social Research," this issue.

First a crucial decision had to be made concerning choice of group activity. Would it be possible to find one single kind of activity sufficiently attractive to 13 year-old boys in all seven countries to establish the necessary degree of cohesiveness in all experimental groups? Without having done any systematic survey of boys' interests in the different countries we decided to make our experimental groups, boys' aviation clubs. This proved to be a lucky choice indeed. On the average 84% of all the boys to whom appeals were made in all the seven countries filled in application sheets for club membership, the range being from 71 to 99 per cent.

So far we succeeded in keeping cohesiveness approximately constant and at a fairly high level in all cultural settings. When trying to evaluate cohesiveness carefully in terms of responses to control scales, however, we were immediately faced with a series of problems that might have been crucial had "cohesiveness" been one of the independent variables in our study. Ordinary measures of cohesiveness based upon responses to questions like: "How much time are you willing to spend in club activity?", "How often would you like your club to meet?" etc., obviously could not be taken at face value, because such things as the total amount of leisure time and the availability of other attractive group memberships varied somewhat from one cultural setting to another. A careful between-nations comparison of cohesiveness therefore seems to presuppose some relative measure, some weighting of absolute scale responses in terms of nation-specific situational factors intervening between the genotype "cohesiveness" and its manifestations in verbal behavior. However, factors mentioned above like amount of leisure time and availability of other group memberships can be thought of as actually influencing the cohesiveness of the experimental groups. We could assume, for instance, that the amount of available leisure time increases or decreases the individual's attraction to a group. In this case the construction of a relative scale would not solve the problem, but we would have to take into account actual differences in cohesiveness depending on situational factors. For example, most of the subjects in England belonged to already existing clubs. The experimenters in England had great difficulties in creating situations in which pressures toward uniformity, and therefore rejection of deviants, would develop. It is conceivable that the existence of alternative group-memberships lowered the cohesiveness of the experimental clubs to such a degree that a successful performance of the whole experiment was endangered.

The alternatives in such a case are twofold: 1.) to use relative scales and assume that situational factors like those mentioned above do not influence the cohesiveness of the group but only modify the verbal responses to the questionnaires (in other words, we question the validity of the measures); and 2.) to assume that such factors have a crucial influence on the general level of cohesiveness of the groups in different countries. In this case we have to vary the group activity from country to country in order to obtain the same level of cohesiveness. This kind of procedure will

be discussed in more detail in connection with the screening of the goal valence variable.

Turning to "goal valence", again our gross impression based upon the control data is that we have been fairly successful in establishing conceptually identical conditions, i.e. by and large identical goals in different settings seem to have aroused approximately the same degree of valence. There is, however, one obvious exception. The goal for all low valence groups was free tickets to a dull documentary movie. In most urban districts in western Europe such a goal is fairly unattractive for 13 year-old boys. In Belgium, however, the subjects were parochial school boys who very seldom had any opportunities to go to the movies. Control measures show that the valence of the film tickets seems to be considerably higher there than in the other countries. In this case we would probably have had more success in establishing conceptually identical situations if we had allowed for more freedom of experimental manipulations, i.e., if different experimental manipulations had been set up to rule out the effects of situational factors. Instead of film tickets as a low valence goal we might in the Belgian groups have chosen something like a dull book as being functionally equivalent to the film tickets in other countries. This example clearly demonstrates that cross-national standardization (in the sense of establishing conceptually identicaal conditions) in some cases can be best achieved at the cost of uniformity of stimulus situations.

Our control measures of the independent variables warrant special comment. Most of our measures were based upon responses to questionnaires. Thus, as one check on the effectiveness of our valence manipulation we used individual responses to the following question (phrased for low-valence groups): "How disappointed would you be if your group did not receive movie tickets?" By checking one of five alternatives (ranging from "I can't imagine anything that would disappoint me more" to "I would not be at all disappointed") each subject was then assumed to indicate indirectly the strength of his desire for the goal. The inferences that can be made, from responses to such a scale, concerning the relative effectiveness in the several experiments of the manipulation of valence are, however, complicated by our lack of comparative semantic knowledge. We know so little about cultural differences in modes of verbal expression that the safest procedure may be to assume that one and the same response reveals the same degree of valence irrespective of the culture of the respondent. We would not be very much astonished, however, if upon systematic explorations in this field we were able to identify cultures in which the subjects by and large had a definitely stronger, more extreme mode of expression than subjects from other cultures. At all events the problems seem worthy of cross-cultural semantical studies. In future comparative research we could then better evaluate the validity of our control measures and—if necessary—make corrections for divergent verbal habits. Such an exploration of culturally conditioned verbal habits seems to be particularly urgent when

because of situational differences we substitute functionally equivalent manipulations for identical stimulus conditions. The functional equivalence must somehow be checked, and this can only be done by relying on verbal responses to control scales.

Measurement of the Dependent Variables

Let us now turn to step (c) in Figure 1, the estimation of the dependent variable (in our case rejection) on the basis of measurements of some behavioral manifestations. Our problem here is analogous to that encountered in the discussion of control measures. In both cases we must examine possible culture-specific "channels" or "ducts" leading from a genotypically defined variable to its overt measurable manifestations. The question is whether expressions of rejection are culturally determined in such a way that with identical instruments we are getting at different aspects of rejection in different cultural settings.

Our measures of manifestations of rejection were based upon two scales, 1) "a working-together scale" designed to measure how much the subjects wanted to have the deviate in their own working sub-group and 2) a "presidential scale" designed to measure how much the subjects wanted the deviate as a president of the club. Table I presents rank order correlations of these two measures in the experiments of six countries.[5]

TABLE I

RANK ORDER CORRELATIONS OF THE DEVIANT'S POSITION ON TWO SOCIOMETRIC SCALES

	Germany	*Belgium*	*Sweden*	*Norway*	*Holland*	*U. K.*
Rho	.80	.56	.63	.66	.49	.62
N	32	24	29	32	50	12
Range of "Working-Together-Scale"	2.50	2.40	2.67	2.50	2.60	1.50
Range of "Presidential Scale"	3.60	2.83	3.00	3.17	3.83	3.60

The estimated mean of the distribution of rho is .63 and the fiducial limits at the 1% level are .48 and .74. With the exception of Germany, the correlations may be considered as samples drawn from the same population. The positive, though on the average not very high correlations, and the relatively small between-nations differences tend to confirm our general expectations concerning multiple "ducts" between genotype and be-

[5] Unfortunately the relevant French data were unavailable at the time this paper was being prepared.

[6] See Edwards, A. L., *Experimental Design in Psychological Statistics*, New York, 1950.

havioral manifestations. Responses to both scales seem to be reflecting somewhat divergent aspects of the same central tendency and each scale can therefore be applied as an imperfect, partial measure of rejection.

The difference between Germany and the other countries is of interest. To estimate the level of significance of the difference we computed a weighted mean value for the five countries and made a test of the significance of the difference between the mean value and the German correlation-coefficient using Fisher's z-transformation.[6] The weighted mean for the 5 countries transformed is 1.009 and the German z-value is .661. The standard error of difference is .205, which gives a critical ratio of 2.14 (p-value of less than .03 for a one-tailed test).

How can this difference be explained? At present we cannot provide more than hunches, some of which may be worth pursuing. We can rather safely assume that in general the correspondence between responses to different sociometric scales is a function of the lifetime of the groups. The less a person knows his fellows, the less differentiated will be his judgments of them. As duration of group membership is held constant in all countries, however, we have to search for other differentiating factors. One hunch leads in the direction of a "true" cultural difference. The difference may be interpreted as indicating a more pronounced tendency to discriminate between *characteristics qualifying for group leadership and characteristics qualifying for personal friendship,* in the five "low-correlation" countries than in Germany.

Using Parsons' concepts[7] we may speak of a lack of discrimination between "particularistic" and "universalistic" role-obligations among the German subjects. Perhaps more thorough-going previous training in group-life and democratic practices had made the subjects in Holland, Belgium, Sweden, England and Norway more sensitive in judgment of interpersonal relationships, in the sense that they are able to pick out differential cues for choices of working mate and group-leader. Such an assumed learned discriminatory power would explain the fact that one and the same person, the deviant, may be strongly rejected as a group leader but at the same time only mildly rejected or even accepted as a working mate—and vice versa. The close correspondence between the two sociometric measures in the German groups would then be interpreted as testifying to a more diffuse, undifferentiated judgment of interpersonal relationships. The cues for friendship and leadership choices may be highly interrelated. There may, thus, be a tendency to choose as friends primarily those who demonstrate leadership abilities *or* a tendency to choose as leaders primarily those who manifest characteristics qualifying for personal friendship.

Summary

In the present paper we have focussed upon some general problems encountered when a set of non-operationally phrased hypotheses are to be

[7] T. Parsons *et. al., Towards a General Theory of Action,* Cambridge, Mass., 1951.

experimentally tested in different cultural settings. We have attempted to show that conceptually identical experimental conditions, in some cases, can best be established by deliberate and systematic deviations from traditional standardization procedures, e.g. by "objectively" different experimental manipulations designed to rule out the effects of specific, culturally determined situational factors intervening between a set of stimuli and a genotypically defined independent variable. Further we have pointed to some of the pitfalls pertaining to evaluation of experimental manipulations in terms of verbal responses to control scales. We suggest semantic studies of verbal behavior in different settings as a means for improving the cross-cultural comparability of control measures, especially important in the case where one has to substitute "functionally equivalent" manipulations for identical stimulus situations. Finally we have commented upon the measurement of the dependent variable, indicating that possible culture-specific "ducts" leading from the genotypical to the phenotypical level may complicate an otherwise simple inference from identical scales to one and the same genotypically defined variable.

Correspondence

We are proud to print the following letter received by Kenneth B. Clark, Issue Author of "Desegregation: An Appraisal of the Evidence," Vol. 9, No. 4, 1953. We are also proud that the initial large printing of this issue is already exhausted. Two thousand more copies are just off the press. With the implications of the Supreme Court decision still to be worked out in practice, the issue is more timely than ever. Have *you* considered giving copies to your school board, your council of social agencies, your professional associates, your friends? Have you considered assigning it to your classes?

M. BREWSTER SMITH
General Editor

August 18, 1954

Dear Mr. Clark:

The Housing Authority of Baltimore City wishes to take this opportunity to advise you of the extreme value secured from your publication, "Desegregation: An Appraisal of the Evidence", which was published in the Journal of Social Issues. The executive staff and the Commission of the Housing Authority of Baltimore City relied heavily upon the material contained in the Journal. Of all of the publications which were used in reaching a decision to recommend a change in policy, your data was by far the most enlightening and the most helpful.

You are to be commended for bringing together in one place the many implications of segregation and for organizing it in such a readable manner. It is our opinion that the booklet should be required reading for anyone, or any group, who is attempting to reach a conclusion relative to the unreasonableness of segregation.

We are pleased to be able to advise you that the document was a most valuable tool to this Agency in reaching its decision to desegregate the entire low-rent housing program in Baltimore City.

Sincerely,

(signed)
ELLIS ASH
Director of Management
Housing Authority of
Baltimore City
Baltimore 2, Md.

To The Editor:

As a charter member of SPSSI, I regret the evasion of large, live, and liberal issues in our journal and the seeming timidity and ideological poverty of our membership.

We used to assume the knowledgeable role of the intelligent participant-observer whether he were academically oriented or not, ideologically pure or not, participating in opposition or in support. We faced issues.

Dean Melby's political sycophancy (see "Academic Freedom in a Climate of Insecurity," 1953, Vol. 9, No. 3) of course has nothing to fear but fear itself. But if one raises significant issues—such as the veracity of Herbert Brownell vs. Harry Dexter White, the influence for democracy of the two men, the trend toward McCarthyism in government and its utilization of religious concepts, the militarization of higher education, and the production and promulgation of synthetic concepts like communist aggression, soviet conspiracy, containment or liberation, and crusade against atheistic communism—then perhaps one has more to fear than fear itself. Deans and Senators perhaps.

I consider the last issue, gingerly readying us for the criminal idea of atomic war, a perversion, invasion, and misuse of our society and its proper notion of "social issues", besides being a waste of good paper since it consists largely of talk about studies rather than report of anything theoretically or practically significant. It is typical of what the military "services" and "their" money are doing to us.

HARRY C. STEINMETZ
Clinical Psychologist
San Diego, California.

INTERNATIONAL SOCIAL SCIENCE BULLETIN

published quarterly by the United Nations Educational, Scientific and Cultural Organization, 19 Avenue Kléber, Paris 16e.

Just issued: Vol. VI, No. 3

PART I: ECONOMIC MOTIVATIONS AND STIMULATIONS IN UNDER-DEVELOPED COUNTRIES

Contributors to this issue: D. Young, G. Balandier, M. J. Herskovitz, R. Firth, D. P. Mukerji, J. H. Boeke, B. F. Hoselitz, A. Doucy, J. Malaurie, J. Matos Mar.

PART II: ORGANIZATION IN THE SOCIAL SCIENCES, REVIEWS AND ANNOUNCEMENTS

Review of Documents, Periodicals and Books—News and Announcements

Annual subscription:
$3.50, 21/-, 1.000 fr.
Per copy: $1.00, 6/-, 300 fr.

U.S. DISTRIBUTOR:
UNESCO PUBLICATIONS SERVICE
475 Fifth Avenue,
New York City, N. Y.

Subscribe Now for 1955!

Please check:
NEW SUBSCRIPTION ☐
RENEWAL ☐
ORDER FOR BACK ISSUES ☐

The Journal of Social Issues
Association Press
291 Broadway, New York 7, N. Y.

Name .....

Street Address .....

City and State .....

☐ Please begin this subscription to the Journal of Social Issues with Volume XI, No. 1. I enclose $4.00 for my subscription to Volume XI (4 numbers).

☐ I enclose .....for subscription for.....years, beginning with Volume XI, No. 1.
(If this is a gift) Please send gift card with donor's name as follows:
.....

☐ Please send me the following previous issues of the *Journal* (Specify Volume and Number): .....
.....

Total Enclosed: $.....

Previous Issues

Yearly subscription (4 issues)	$4.00
Group subscriptions (to single address)	
10 or more	2.50
20 or more	2.00
Back issues (except Vo. VII, Nos. 1 & 2)	
single copy	$ 1.25
per dozen	10.00
twenty-five copies	17.50
Vol. VII, Nos. 1 & 2: Cloth binding, per copy	3.00
Supplement issues	.50

Volume I: 1945
- No. 4 The Psychological Problems of Bureaucracy. Goodwin Watson, Issue Editor.

Volume II: 1946
- No. 2 Measuring Public Attitudes. Angus Campbell, Issue Editor.
- No. 3 American Military Occupation: A Test for Democracy. Waldemar A. Neilson, Issue Editor.
- No. 4 Action and Research: A challenge. David Krech, Issue Editor.

Volume III: 1947
- No. 1 Toward a Common Ground—International Social Science. Jerome S. Bruner, Issue Editor.
- No. 3 The Mass Media: Content, Function, and Measurement. Franklin Fearing, Issue Editor.
- No. 4 Social Research in Political Decision. L. M. Hanks, Jr., Issue Editor.

Volume IV: 1948
- No. 1 Citizen Participation in World Affairs: Problems and Possibilities. Eleanor Macoby and Ben Willerman, Issue Editors.
- No. 2 The Dynamics of the Discussion Group. Leland Bradford and J. R. P. French, Issue Editors.
- No. 4 The Social Psychology of Physical Disability. Lee Myerson, Issue Editor.

Volume V: 1949
- No. 2 Community Self-Surveys: An Approach to Social Change. Claire Selltiz and Margot Haas Wormser, Issue Editors.
- No. 3 Consistency and Inconsistency in Intergroup Relations. Isidor Chein, Morton Deutsch, Herbert Hyman, and Marie Jahoda, Issue Editors.
- No. 4 Social Values and Personality Development. Max L. Hutt and D. R. Miller, Issue Editors.

Volume VI: 1950
- No. 1 Community Service and Social Research—Group Psychotherapy in a Church Program, Hubert Coffey, Mervin Freedman, Timothy Leary, Abel Ossorio, Issue Authors.
- No. 2 Educational Dynamics: Theory and Research. Herbert A. Thelen, Issue Author.
- No. 3 Professional Women in Modern Society,May V. Seagoe, Issue Editor.
- No. 4 Values and the Social Scientist. Kenneth Benne and G. E. Swanson, Issue Editors.

Volume VII: 1951
- Nos. 1 &2 Social Policy and Social Research in Housing. Robert K. Merton, Patricia Salter West, Marie Jahoda, and Hanan C. Selvin, Issue Editors.
- No. 3 Human Relations Research in Large Organizations. E. Jacobson, R. Kahn, F. Mann, and N. Morse, Issue Editors.
- No. 4 American Culture and Personality. Abraham H. Maslow, Issue Editor.

Volume VIII: 1952
- No. 1 Intergroup Contact and Racial Attitudes. John Harding, Issue Editor.
- No. 2 Group Methods in Psychotherapy, Social Work, and Adult Education. Harry Lerner and Herbert C. Kelman, Issue Editors.
- No. 3 Social Relations Skills in Field Research. Floyd Mann and Ronald Lippitt, Issue Editors.
- No. 4 Sociocultural Approaches to Medical Care. Kenneth Herrold and Dorothy Lee, Issue Editors.

Volume IX: 1953
- No. 1 Trade Unions and Minority Problems. Daniel Bell and S. M. Lipset, Issue Editors.
- No. 2 Community Change: An Action Program in Puerto Rico. Charles F. Cannell, Fred G. Wale, and Stephen B. Withey, Issue Editors.
- No. 3 Academic Freedom in a Climate of Insecurity. Ernest O. Melby and M. Brewster Smith, Issue Editors.
- No. 4 Desegration: An Appraisal of the Evidence. Kenneth B. Clark, Issue Author.

Volume X: 1954
- No. 1 Human Problems in the Changing South. Arthur J. Bachrach and Gordon W. Blackwell, Issue Editors.
- No. 2 Human Relations Training for School Administrators. Arthur W. Combs and Robert S. Fisk, Issue Editors.
- No. 3 Human Behavior in Disaster: A New Field of Social Research. Dwight W. Chapman, Issue Editors.

Supplement Series
- No. 1 Social Responsibility. George Brock Chisholm.
- No. 2 Social Science and the Atomic Crisis. Hornell Hart.
- No. 4 Prejudice: A Problem in Psychological and Social Causation. Gordon W. Allport.
- No. 5 Some Aspects of Social Process. A. T. M. Wilson.
- No. 6 Psychological Impediments to Effective International Cooperation. Gunnar Myrdal.
- No. 7 Human Potentialities. Gardner Murphy.
- No. 8 Cultural Discontinuities and Personality Transformation. Margaret Mead.